# AMERICAN JOURNAL of PHYSICS

*A Journal Devoted to the Instructional and Cultural Aspects of Physical Science*

VOLUME 15, NUMBER 2 — MARCH–APRIL, 1947


## Survey of Proximity Fuze Development


CHESTER H. PAGE AND ALLEN V. ASTIN

*Ordnance Development Division, National Bureau of Standards, Washington 25, District of Columbia*


This paper was prepared in response to a request from the editors of the *American Journal of Physics*. The authors are associated with the Ordnance Development Division of the National Bureau of Standards, which group served during the war as the central laboratories for Division 4, NDRC, in the development of proximity fuzes for fin-stabilized missiles. Fuzes for spin-stabilized missiles were developed by Section T of OSRD, with central laboratories at the Applied Physics Laboratory of Johns Hopkins University. The authors are indebted to J. Allen Hynek for information supplied concerning the fuze program for spin-stabilized projectiles.

Historical aspects of the proximity fuze program are adequately covered in:


1. J. P. Baxter, "Proximity fuzes: a challenge to air power," *Atlantic Mo.* p. 75 (Sept. 1946).


A review of the operational use from a tactical point of view is given in:


2. Col. H. S. Morton, "The VT fuze, an outstanding weapon," *Army Ordnance* p. 43 (Jan.–Feb. 1946).


Technical aspects of various types of fuze are covered in the following articles:


3. F. Rocket, "Proximity fuze," *Electronics* pp. 110–111 (Nov. 1945).
4. R. D. Huntoon and B. J. Miller, "Generator-powered proximity fuze," *Electronics* p. 98 (Dec. 1945).
5. H. Selvidge, "Proximity fuzes for artillery," *Electronics* p. 104 (Feb. 1946).
6. R. G. Peters, "The radio proximity fuze," *Communications* p. 45 (Oct. 1945).
7. H. Diamond, "The radio proximity fuze," *National Radio News* p. 16 (Dec.–Jan. 1945–1946).
8. C. Kleiderer, "Proximity fuze taught new techniques," *Modern Plastics* p. 133 (Nov. 1945).
9. W. S. Hinman, Jr., and C. Brunetti, "Radio proximity fuze design," *NBS Research* **37**, 1 (July 1946).


The literature references (superscript numerals) in the text that follows pertain to the correspondingly numbered items in the foregoing bibliography.

It should be emphasized that many features of proximity fuze design are still military secrets, and the scope of this article is limited by the rules of military security. Its publication is approved by the Army Ordnance Department, as well as by the National Bureau of Standards.

---

PROXIMITY fuzes are the realization of a long cherished goal in the field of modern ordnance,[2] since such fuzes detonate missiles automatically upon approach to a target and in a position to inflict maximum damage. Proximity fuzes were developed during the war for use on a wide range of missiles—antiaircraft and artillery projectiles, bombs, rockets, and trench mortar shells. Their operational use, under the code name of VT fuzes, was one of the most closely guarded secrets of the war, and only after the complete defeat of our enemies was it possible to disclose the important role these fuzes played in our victory.[1–9] The development and production of VT fuzes was an outstanding illustration of American technological supremacy; not only did we lead our allies in this novel field, but we were also years ahead of Germany and Japan in this phase of modern warfare.

Although the rapid growth of enemy air power provided the main impetus for the development of proximity fuzes, the antiaircraft use is only one of many important applications. As an air defense weapon, the value of the proximity fuze

is easily appreciated. It is not necessary to hit an airplane to bring it down; the only requirement is that the projectile approach the airplane within the lethal operating range of the fuze, usually about 60 or 70 ft. This means that the effective size of the target is increased many fold—of the order of 20 to 50 times for the average airplane. The increase in effective target area represents roughly the advantage, in long-range antiaircraft fire, of VT fuzing over contact fuzing.

In fairness, however, VT fuzes should be compared to time fuzes which had been developed previously to detonate after a selected time of flight, presumably in the vicinity of the target. To use a time fuze effectively, it is necessary to determine accurately and quickly the distance to the target and to set the fuzes accordingly just before firing. Modern radar ranging technics improved appreciably the effectiveness of fire with time fuzes; but even with the most accurate ranging methods, the time fuze was only about one-fourth as effective as the proximity fuze. Further advantages of the proximity fuzes in antiaircraft fire are: (i) all bursts on the target are at approximately the proper point on the trajectory to inflict maximum damage; (ii) no setting or adjustment of the fuze before firing is necessary; it is a completely automatic device; and (iii) the fuze provides a considerable surprise factor because the pilot of the target aircraft usually is not aware that he is being shot at until he is hit.

The next major application of VT fuzes, and the one for which the largest number of fuzes were produced, is the "air burst" of a missile on approach to a ground target. The advantage of an air burst had been appreciated in artillery circles for many years, and its superiority over the more usual surface bursts against many types of targets has been conclusively demonstrated in comparative tests. The explanation of the superiority is a matter of simple geometry. Fragments from an exploded artillery shell, bomb or other missile travel in essentially straight lines throughout their lethal range. Thus, when a missile bursts on the surface of the ground, its fragments will damage only those targets that are practically in line of sight with the point of impact. A ditch, foxhole, or even a slight bump on the ground offers a high degree of protection to a soldier from everything except a direct hit. Similarly, inanimate targets such as guns, trucks, and parked aircraft can be protected from projectiles exploding on the ground by pits, revetments, and sand bag barriers, as well as by each other. If, on the other hand, the missile explodes in the air, many of its fragments will be hurled down into foxholes and revetments, and behind protecting walls, so that the shielding effect of these supposedly safe locations is nullified. In essence, an air burst increases the effective size of the target (as in antiaircraft fire) since a direct hit is not necessary to inflict lethal damage. The height at which an air burst produces its maximum damage is rather critical; it must be high enough to expose an appreciable area of targets but not so high that the targets are beyond the lethal range of the fragments. It has been established that an air burst in the optimum height range is 5 to 20 times as effective as a surface burst against various types of protected targets, the superiority factor varying with the missile, the nature of the target, and its degree of protection. This means that a single salvo of VT-fuzed artillery shells or a single bomber load of VT-fuzed bombs can be as effective as 5 to 20 salvos or bomber loads fuzed in the conventional manner.

Before proximity fuzes were available, attempts were made to secure air bursts by means of time fuzes. Although reliable performance was obtained in very short-range artillery fire, the burst locations were entirely too erratic in long-range fire and in bombing operations; and, in the latter case, the advantage of the air burst was largely neutralized by the tremendous variation in burst heights. With the proximity fuzes, however, the full advantage of the air burst is obtained since the fuzes are designed to function automatically at the proper position above the ground. Furthermore, this performance is achieved under a variety of conditions where accurate time fire would be impossible, and without the complications of estimating the "time to target" and then of setting the fuze before firing from a gun or release from a bomber.

Examples of other applications of the VT fuze are the production of air bursts in large incendiary or fire bombs to enhance the spread of fire by avoiding the loss in the crater formed by a

contact explosion, and the production of accurately located air bursts of large bombs of the "block-buster" type to increase the area of destruction due to blast effect.

Extreme security precautions accompanied the development and use of proximity fuzes, and every possible effort was taken to keep the enemy from learning that we had succeeded in producing such a weapon. The major reason for these unusual precautions was of course that we did not want to assist the enemy in any way in his own efforts to produce a proximity fuze. So potent is the proximity fuze as an air defense weapon that had the enemy had these fuzes in any appreciable quantity, it would have caused a major setback in our own air offensive. Both the strategy

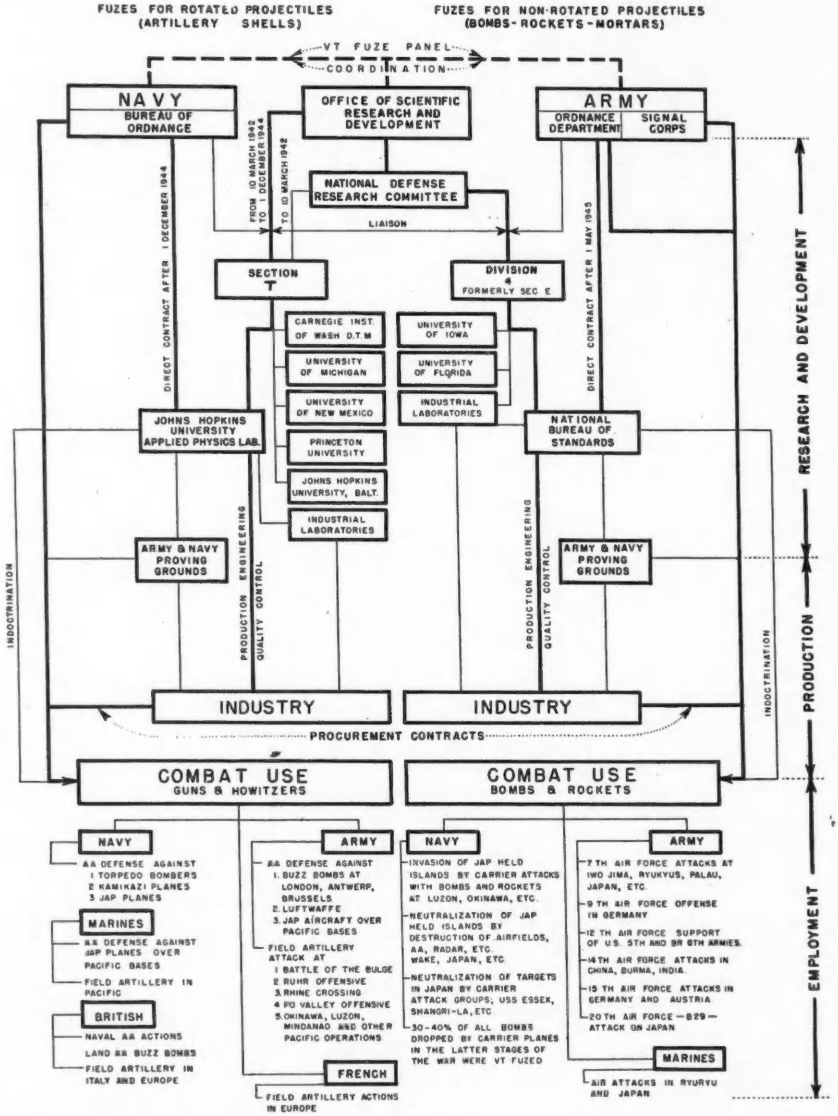



FIG. 1. Wartime organization chart of the proximity fuze project. [Based on a chart prepared under the direction of Capt. J. P. Teas, of the VT Fuze Detachment.]

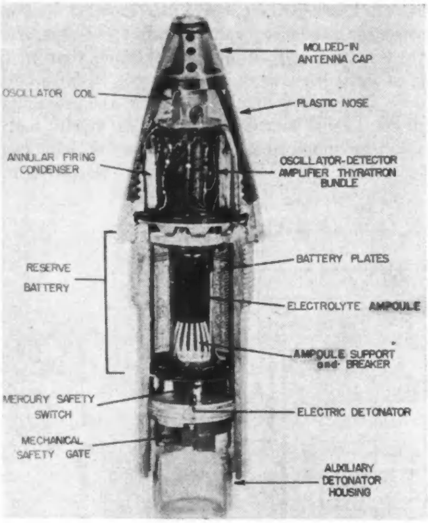



FIG. 2. View of sectioned model of a typical radio proximity fuze for a shell, showing the principal components of the fuze. [Applied Physics Laboratory photograph.]

and the tactics of our plans of air bombardment would have been seriously affected. Fortunately, the enemy, as we learned after the end of the war, were a long way from having successful proximity fuzes, although they were working vigorously to that end. To accomplish our objective of keeping the enemy uninformed of our own achievement, the extent of our use of proximity fuzes was severely restricted. Until December 1944, when the VT fuze was used effectively to help stop Von Runstedt's counteroffensive, no VT fuzes were used when there was any possibility that even one might be recovered by the enemy. From January 1943, when the first VT fuze was used against the enemy in the Pacific, until December 1944, the use of the fuzes was confined to the antiaircraft role and to localities where unexploded shells would fall either in the deep waters of the ocean or in completely protected territory. Even our battleships were forbidden to use the VT fuzes against attacking bombers when the ships approached too close to enemy-held islands.

It is not intended to survey the operational use of proximity fuzes in this paper since that has been done excellently elsewhere;[1,2] however, a brief résumé of such use may be seen in the chart of Fig. 1.

### General Method of Operation

A VT fuze consists essentially of the following parts: (i) a radio transmitter and receiver, (ii) a selective amplifier, (iii) an electronic switch, (iv) a detonator, (v) an electric power supply, and (vi) arming and safety devices. The radio transmitter sends out waves which, when they strike a target such as an airplane or the ground, are reflected back to the fuze. The reflected waves differ in frequency from the transmitted waves owing to the relative velocity of the fuze and target (Doppler effect). The radio receiver in the fuze picks up the reflected waves and detects the difference frequency. The latter is amplified, and when a certain level of intensity is reached, depending primarily on position with respect to the target, the amplified signal trips an electronic switch or thyratron. This action fires the detonator, and the projectile is exploded. The arming system of the fuze insures that the fuze will be inoperable until it is at a safe distance from the gun from which it is fired or from the bomber from which it is dropped. Electric energy must be provided to operate the radio transmitter and receiver, and all VT fuzes have tiny built-in power supplies for that purpose.

The foregoing outline of operation covers the essential features of practically all VT fuzes. The engineering and design features, however, differ appreciably in various types of fuze, depending primarily on the intended application. A major separation of responsibility for development and production was effected early in the program in order to expedite the work. A basis for the division of responsibility was provided by the fundamental difference in the two major classes of ordnance missiles, which relates to the method of stabilizing the missile in flight. Stabilization can be effected either by spinning the projectile rapidly, as with artillery shells, or by providing fins on the missile, as with bombs, some rockets, and trench mortar shells. Since the design of a VT fuze must be adapted to the type of projectile stabilization, the division of responsibility in the VT fuze program quite naturally resolved itself into fuzes for rotated and fuzes for nonrotated projectiles. Section T of OSRD undertook initial proximity fuze development in the United States, but early in 1941 concentrated on the development of VT fuzes for rotated projectiles, with the Navy Department in charge of production. Division 4 of NDRC, OSRD, took over the develop-

ment of fuzes for nonrotating missiles, and the War Department handled the production.*

A major design consideration for the fuzes for rotated projectiles was ruggedness. All parts of the fuze had to withstand the terrific impact of being fired from a gun, thereby being subjected to accelerations up to 20,000 times that due to gravity. This was a particularly critical requirement for the miniature vacuum tubes of the radio set, but after an intensive development program, satisfactory tubes were designed and produced in enormous quantities. An indication of the significance of this achievement is obtained from the record of the German attempts to build proximity fuzes. So insurmountable did they consider the problems of designing a radio tube which could be fired from a gun that they decided the outlook was hopeless and never tried it, although they attempted all sorts of other proximity devices which ultimately proved inadequate.

The safety and arming features of the fuze for rotated projectiles depend on rotation of the fuze for operation. Thus, unless the fuze undergoes the combined shock of propulsion and sustained rotation which is encountered in the gun barrel, it will remain inoperative. This means that dropping or other rough handling cannot set off the fuzes; ammunition handling and loading procedures involve no hazard. One of the safety features is the battery that supplies the power to run the radio set and the electronic switch. The battery is of the wet type but is dead until it is fired in a gun. When fired, the battery is activated and electric energy becomes available to operate the fuze. A photograph of a sectionalized typical fuze is shown in Fig. 2.

The VT fuzes for bombs and rockets did not have to meet the extreme requirement for ruggedness necessary for the shell fuzes. However, stability was a very important factor since the fuze had to remain operative while undergoing the intense vibration produced in a bomb when dropped from high altitude. Another major critical requirement for the bomb and airborne rocket fuzes was that of temperature. The airplanes which carried the weapons operated at very high altitudes, where temperatures as low as −40°C are not uncommon. Dry batteries and even batteries of the type used in the shell fuze are inoperative at these low temperatures. This meant that VT fuzes which used battery-type power supplies could not be employed for high altitude bombing operations unless special precautions were taken to keep the fuzes warm. This proved to be a very unsatisfactory solution; so a new type of power supply was developed, the wind-driven generator. All VT fuzes for bombs and fin-stabilized rockets now have this type of power supply. The generator-type power supply has a further advantage of safety. The fuze is inoperative, owing to lack of electric energy until it is in a fast moving air stream. An additional safety feature is that the fuze remains unarmed until the generator has completed a predetermined number of rotations, thus ensuring that before the bomb or rocket equipped with the fuze becomes operable it will be well away from the aircraft that releases it. A photograph of a typical bomb fuze, sectionalized, is shown in Fig. 3.

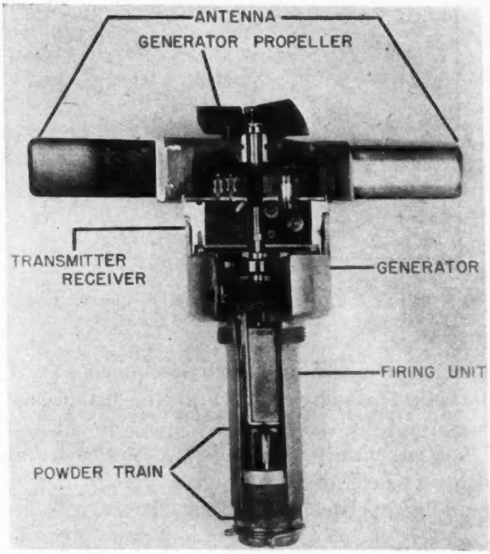



FIG. 3. Sectioned model of a radio proximity fuze for a bomb, showing the principal components.

## Oscillator Design

The oscillator is the heart of the radio proximity fuze in that it is the device that actually

* Merle A. Tuve was Chairman of Section T, and L. A. Hafstad, his Chief Deputy. Alexander Ellett was Chief of Division 4, and Harry Diamond was in charge of the Division's Central Laboratories at the National Bureau of Standards.

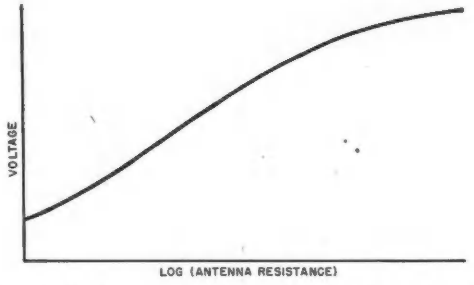



FIG. 4. Typical loading curve of an oscillator, showing voltage as a function of load resistance.

detects the target and develops a signal voltage, the value of which depends upon the distance and velocity of the target.

Any electrically reflecting object in the electromagnetic field of the fuze transmitter reflects some energy back to the fuze. If the reflected wave is in such a phase as to induce an opposing current in the transmitting antenna, the net antenna load is less than if there is no reflection. Hence, the transmitting oscillator is presented with a higher radiation resistance. Conversely, when the reflecting object (target) comes a quarter wavelength closer, the antenna current is larger than normal; this is equivalent to a reduced resistance. Continuous motion of the target can be interpreted as a cyclic variation of the complex radiation impedance of the fuze antenna. For practical purposes, the effect upon the oscillator is that of a sinusoidally varying resistive load since the reactive component of the actual variation has little effect. The behavior of the fuze oscillator can therefore be expressed in terms of a loading characteristic, or a curve of some reference voltage as a function of radiofrequency load resistance.

An idealized oscillator acts as a normal generator—a constant voltage source with a constant series impedance, or a constant current source with a shunt impedance. In the idealized case, which is fairly well approximated in practical fuze designs, the reference voltage is proportional to the terminal (radiofrequency) voltage appearing across the load. Analysis of this situation leads to a loading curve that is a smooth **S** curve (Fig. 4) when the voltage is plotted as a function of the logarithm of the load resistance. The reason for using the logarithm of the antenna resistance $R$ is that different geometric arrangements of the antenna members can give different levels of operating resistance, but a given target at a given distance will produce the same *percentage* change of resistance in each case.

The effectiveness with which the oscillator converts target presence into signal voltage is obviously measured by the slope of the loading curve. Hence, we define the oscillator sensitivity $S$ by the equation

$$S=\frac{\mathrm{d}V}{\mathrm{d}\ln R}=\frac{\mathrm{d}V}{(\mathrm{d}R)/R};$$

$S$ is measured in volts per fractional change of load.

The actual signal voltage (variation of the reference voltage) is very small under normal conditions and must be amplified before it can be used to trigger a thyratron tube, which acts as a sensitive relay to fire the detonator and bring about explosion of the bomb. The signal voltage required for operation is a very small fraction of the reference voltage, which implies that any microphonic noise or other very small spurious variation of the reference voltage can produce a false signal. This indicates the need for careful workmanship and design to insure rigidity in the oscillator assembly and in the internal structure of the oscillating vacuum tubes.

## Amplifier

The amplifier, usually a single stage using a subminiature pentode, magnifies the tiny signal voltage to a usable level but allows the incorporation of frequency selective response. Proper frequency selectivity not only insures reliable functioning but also offers a means of "burst control." The type of response curve depends to a great extent upon the radio sensitivity pattern of the fuze; the main differences in patterns are associated with the type of antenna used, and other details of circuit construction.

Some bomb fuzes (Fig. 3) used a transverse antenna consisting of a center-fed dipole mounted transversely to the bomb axis. Such antennas had the sensitivity pattern indicated in Fig. 5($b$). All shell and rocket models and many bomb models used the body of the projectile itself as the

antenna. This mode of transmission is referred to as longitudinal excitation, in contrast to transverse excitation. Antenna current was excited by means of an insulated cap, or ring, at the forward end of the fuze, connected to a high-voltage point of the oscillator (Fig. 2). The relative amounts of metal fore and aft of the insulator control the driving point resistance, and hence the operating point on the loading curve. The sensitivity pattern is essentially as shown in Fig. 5(*a*).

It is apparent from Fig. 5(*a*) that the radiation of a transversely excited fuze is maximum along the bomb axis and varies slowly with deviation from this direction. Hence, bombs dropped from different altitudes and having corresponding angles of approach to the ground will have approximately equal signal voltages at equal heights above ground. In general, however, the bombs falling from higher altitudes will have larger vertical components of velocity, hence signal voltages of higher frequency. If height of burst is to be independent of altitude of release, a band-pass amplifier is indicated. The gain over the band should be such that the signal voltage at the desired height is just sufficient after amplification to trigger the thyratron. A typical amplifier response curve, obtainable with relatively simple and compact circuits, is shown in Fig. 6(*b*).

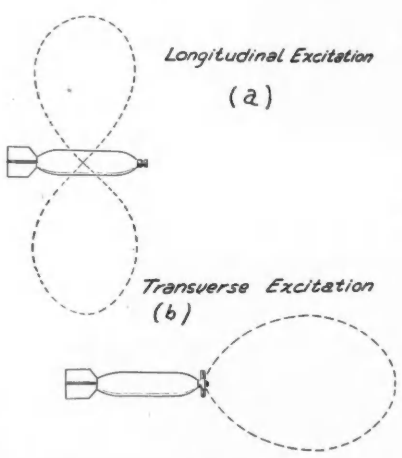



FIG. 5. Directional sensitivity patterns for radio proximity fuzes: (*a*) for a longitudinally excited fuze (one that uses the missile as an antenna); (*b*) for a transversely excited fuze (one that has its own antenna mounted at right angles to the axis of the missile).

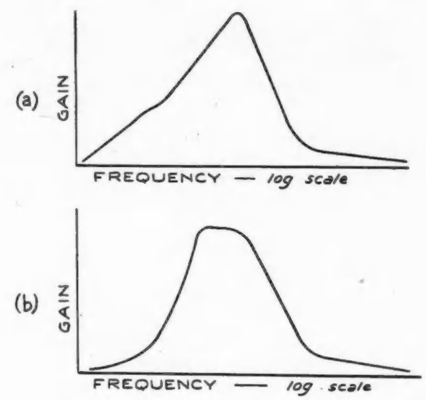



FIG. 6. Typical amplifier response curves for radio proximity fuzes: (*a*) for use with a longitudinally excited fuze; (*b*) for a transversely excited fuze.

On the other hand, the sensitivity pattern of longitudinally excited fuzes is such that the signal is weaker the steeper the missile trajectory. But steeper trajectories are usually associated with higher vertical approach velocities, hence higher signal frequencies. Thus, an amplifier response curve in which the gain increases with frequency up to a maximum value is desired [Fig. 6(*a*)]. Above peak frequency, the gain should be reduced to a low value.

The rapid amplifier cut-off above operating frequency range is desirable to minimize spurious signals due to oscillator vibration or other causes. By proper design and sufficiently rigid mounting of components, all natural microphonic frequencies are kept high enough to be strongly attenuated by the amplifier. (The gain is actually less than unity for frequencies considerably above the Doppler range.) In generator-powered fuzes, a potential source of the vibration signals is in the rotating members of the power supply. The obvious requirement that the fundamental frequency of the rotating system be sufficiently attenuated places a lower limit upon generator speeds.

Another phase of amplifier design was of primary importance for shell and rocket fuzes.[4,5] It is associated with operation against airborne targets which, as targets, differ from the earth's surface in several important respects. For design purposes, the major difference is that fuzes are

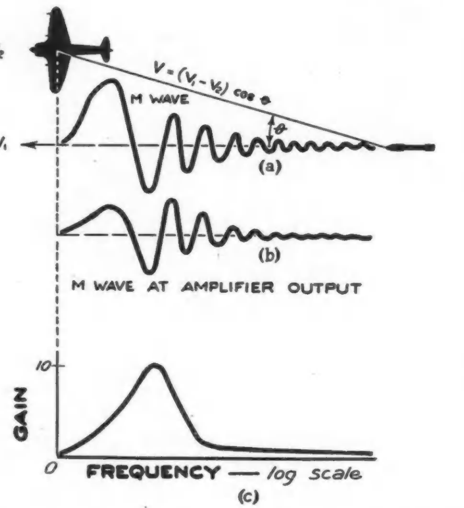



FIG. 7. Diagrams showing modulation ($M$-wave) produced in a radio proximity fuze when approaching an airborne target.

employed against aircraft in "passing shots" rather than "approaching shots." A fuze passing a localized target generates a signal voltage, the phase of which depends upon the distance (in wavelengths) to the target. The amplitude of the signal increases with nearness to the target. As the projectile reaches its nearest point of approach to the target, the rate of change of distance to the target becomes zero. This reduction of instantaneous frequency—a stretching of the phase scale—is indicated in Fig. 7, which presents the so-called $M$-wave showing the instantaneous signal voltage at points along the trajectory. As the projectile travels along the indicated path with constant speed, the signal voltage as a function of time follows the same curve, and the space rate of phase change becomes the instantaneous Doppler frequency, following the familiar relation $F=2V\cos\theta/\lambda$, where $V$ is the velocity of the missile, measured in a coordinate system in which the target is at rest, $\theta$ is the angle between the velocity vector and the line joining the missile and target, and $\lambda$ is the wavelength of the radiation from the source.

The use of a peaked amplifier [Fig. 6($a$)] favors the $M$-wave in the region of appropriate relative velocity between the projectile and plane, so that the $M$-wave as seen at the amplifier output appears as in Fig. 7($b$). This results in shell burst at a certain "lead angle" ahead of nearest approach; the design lead angle is chosen in terms of the geometry resulting from the combination of shell velocity and fragment velocity, to maximize the number of fragments hitting the target.

### Target Reflection and Operating Height

The effect of a perfectly conducting sheet of infinite extent, insofar as the reflection of electromagnetic waves is concerned, can be represented in terms of a mirror image of the source. Thus, a bomb fuze at height $h$ above ground is effectively subject to the radiation from an identical fuze at a distance $2h$. The fractional change of effective radiation resistance is obviously equal to the ratio of received antenna current (due to the field of the image) and original transmitting antenna current. The received antenna current is obviously proportional to the radiation from the image toward the fuze and to the reception pattern of the fuze in the direction of the image. Each of these terms means the same thing—the radiation pattern, $f(\theta)$, of the fuze. The received current is also inversely proportional to the distance between the "source" and the receiver. To express properly the ratio between this received current and the total antenna current, we must essentially normalize the radiation pattern $f(\theta)$ by dividing by the total power radiated in terms of $f(\theta)$ integrated over a surrounding surface. By virtue of axial symmetry (longitudinal excitation), the final result is simple; it is

$$\Delta R/R \propto G(\theta)/h,$$

where $G(\theta)[\equiv f^2(\theta)/2\pi\int_0^\pi f^2(\theta)\sin\theta d\theta]$ is the power gain of the antenna in the desired direction, compared to an isotropic radiator of the same total power.

Since the actual earth is not a perfect reflector, this ideal response must be reduced by a factor $N[<1]$ representing the reflection coefficient of the earth's surface in the target region. Incorporating the sensitivity $S$ of the oscillator, previously discussed, we find the amplitude of the signal voltage to be given by

$$V \propto SNG(\theta)/h,$$

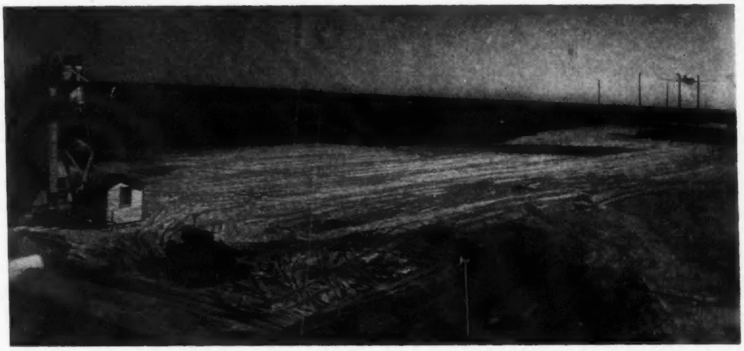


FIG. 8. Experimental range at Fort Fisher, North Carolina, showing firing tower (on the left) and a mock aircraft target made of chicken wire (on the right). A proximity burst may be seen just above the target.

FIG. 9. Target arrangement at the New Mexico Proving Ground, showing an airplane suspended between two towers. [Applied Physics Laboratory photograph.]

and the proportionality constant is readily evaluated by use of radiation theory.

Since the phase of the signal voltage changes periodically with every wavelength change in the distance between the fuze and its image, and the fuze approaches its image with speed $2v$, the frequency of the signal voltage is given by

$$F=2v/\lambda,$$

where $v$ is the vertical component of the bomb's velocity. Thus the whole fuze operation can be considered as transmission of a radiofrequency signal, reception of a reflected signal of higher frequency, detection of the Doppler difference frequency, and amplification of the Doppler frequency.

The equation for signal voltage as an inverse function of height can be solved for the height at which a predetermined signal level will be reached. The predetermined signal level for fuze function depends upon Doppler frequency by way of the amplifier response curve, and so upon bomb velocity. But the vertical striking speed and striking angle are dependent on firing or release conditions and can be determined from ballistic tables for both shells and bombs; hence, we can compute appropriate burst heights if the sensitivity pattern is known.

The sensitivity pattern of any bomb and fuze combination is readily found by supplying the fuze oscillator with a modulated supply voltage and measuring field strength with a simple receiver. If the bomb is mounted on a calibrated turntable in an open space, a complete pattern can be taken. Absolute values of field strength are not needed.

Reflection from an airplane is not a simple phenomenon, either in theory or in practice. At large distances, the target looks small and reflects a signal that is inversely proportional to the square of the distance. At very close range, the spread of a modern bomber makes it almost equivalent to an infinite plane, so that the reflection law becomes that of the inverse first power. The useful range of proximity-fuzed shells and rockets is in the transition region. Superposed upon this effect is the directional response of an airplane; the airplane, excited by induced radio-frequency currents, is a complicated antenna with a correspondingly complicated directivity pattern.

The practical compromise viewpoint is to consider an airplane as reflecting $n$ times as much signal, on the average, as a resonant half-wave doublet. The range of variation of $n$, the number of doublets equivalent to the plane, can be found roughly by function tests of calibrated fuzes. This evaluation is useful for design purposes.

Actual design tests, as well as service tests, of fuzed projectiles against aircraft were made either by shooting projectiles past a full size mock-up craft of chicken-wire construction strung from poles (Fig. 8) or against an actual airplane suspended between high towers (Fig. 9). Three-dimensional location of bursts was accomplished visually and photographically from widely spaced observation posts. The results of firing many rounds of ammunition can be expressed in terms of a "burst surface," an imaginary surface that detonates fuzes by "contact."

### Power Supply

The requirements on the electric power supply for proximity fuzes are quite severe. The first obvious limitation is upon physical size; the available space is small. Ordinary dry cells will deliver sufficient total energy per unit volume to satisfy the total energy requirements. This implies that scaled down dry cells can be made of such a size as to supply the needed power at the needed voltages for the needed length of time, which is usually short. Dry cells, however, do not keep indefinitely in storage, particularly in tropical climates. This restriction could be overcome by constructing fuzes so as to allow batteries to be plugged in, and having suitable storage depots of fresh batteries. Under certain conditions of use, this method might be feasible. General use, on the other hand, requires fuzes to operate properly at temperatures of −40°C, as encountered in the bomb bays of airplanes at high altitude, and 60°C, as can be encountered in tropical warfare. Dry cells will not operate at subzero temperatures and will deteriorate rapidly in the tropics.

One way out of the difficulty is to use a wet cell battery, with an electrolyte that is operable throughout the temperature range. Such electrolytes are available but produce extremely rapid local action and battery deterioration. With a liquid, however, it is possible to keep the cell dry until immediately before use. This was done successfully in the shell fuzes, where high-speed rotation (of the shell) and its accompanying centrifugal force were present. The battery was layer-built of annular cells (Fig. 2); a glass ampoule of electrolyte was properly mounted in the center of the resulting "pipe" structure. Under the firing acceleration in the gun, the ampoule was crushed against a base block and the freed electrolyte distributed by the spin of the shell.[5]

It was important that the design of the battery be such that activation was completed within fairly well defined and small time limits. Another major practical problem with the reserve battery was the design of the glass ampoule. It had to be sufficiently strong to withstand severe shocks from rough handling, yet weak enough to break properly when fired in a gun. Since the two requirements are essentially in opposition, considerable experimentation was necessary before both could be satisfied.

This solution to the problem is not suitable for nonspinning projectiles, such as bombs. The development of power supplies for bomb, mortar, and rocket fuzes accordingly followed a different

approach. A series of wind-driven alternators was developed, some driven by external Bakelite windmills and some by enclosed, metal high-speed air turbines. They were all similar in size and general design.

The alternator comprised a fixed stator surrounding a permanently magnetized rotor. The rotors were of Alnico, approximately $\frac{1}{4}$ in. thick and 1 in. in diameter. A low-voltage winding of a few turns supplied filament voltage for the various tubes; a high-voltage winding of many turns supplied alternating current to be rectified and filtered for plate supply of the oscillator, amplifier and thyratron.

Since the emf induced in a coil is proportional to the rate of change of flux through the coil, and since the flux is derived from a permanent magnet, the open circuit voltage of the generator is proportional to the angular speed of the magnetized rotor. This, in turn, increases with bomb or rocket speed through the air. The proximity fuze, however, needs supply voltages that stay within a relatively narrow range, in order to assure proper operation of the electron tubes. Simple alternators of the type described have considerable inductance. Furthermore, the generated frequency is proportional to the rotor speed. If the alternator feeds a resistive load, the magnitude of which is small compared to the inductive impedance of the winding, the output current (hence, also load voltage) will be substantially independent of speed. This follows because the proportionality of emf to rotor speed exactly compensates the proportionality of inductive impedance to frequency. Thus, for any resistive load, sufficiently high speeds will make the alternator inductance override the load resistance, and the output voltage will become constant. The practical problem is to achieve a flattening off of the voltage-speed curve at lower speeds. Voltage regulation can be acheived by tuning with a capacitance either in series or in shunt with the load.

In practice, a shunt condenser was used more than the series resonant condenser. The shunt condenser was connected to a dummy load. Since the reactance of a capacitor decreases with increasing frequency, the generator "sees" a heavier load at higher speeds. This increase of load can easily be adjusted to compensate the increase of open-circuit voltage of the generator at higher speeds. Parameter adjustments were found to be less critical than with the series resonance method.

Each of the two stator windings is excited by a combination of normal rotor flux and flux induced by the load current in the other winding. The coupling between the two windings is readily adjusted (in design, not individually) so that good regulation of the high voltage results in good regulation of the low voltage.

Selenium cells, about $\frac{1}{4}$ in. in diameter and 1/25 in. thick, were used as rectifiers for the plate power supply. They had sheet aluminum or iron for a base; selenium was coated on the base by any of various methods; the selenium surface was oxidized and heat treated, and finally the other terminal was provided by spraying on a thin coating of low-melting-point alloy. Such cells were stacked to a number compatible with the voltage limit per cell. Light spring pressure maintained steady contact between cells. Further details of the rectifier and regulator for generator-powered fuzes are given in reference 9.

### Safety and Arming

Since proximity fuzes are, by their very nature, sensitive to their surroundings, it is most important that they be kept inoperative until the missiles which carry them are well away from the firing or launching point. The act of making a fuze operative so that it will detonate the missile when a triggering signal is received is called *arming*. Arming in all proximity fuzes consists of essentially three stages: (i) activation of the power supply so that the vacuum tubes of the fuze become operative; (ii) connection of the electric detonator to its proper place in the fuze circuit; and (iii) removal of a mechanical barrier between the detonator and the booster. Completion of all three stages is necessary before the missile can be detonated. If the first stage only is completed, a target signal may trigger the thyratron but neither the electric detonator nor the missile will explode. If the first and second stages only are completed and the detonator fired, the booster will be safe because a mechanical barrier prevents the small explosive shock from the detonator from reaching the booster. The third and last stage of arming removes the me-

chanical barrier. Location of the arming elements in typical fuzes is indicated in Figs. 2 and 3.

The methods of controlling the arming stages vary with the different types of fuze and are related to the ballistic properties of the missiles for which the type is intended. In shell fuzes, the first arming stage—activation of the reserve battery—requires the combined setback (acceleration) and spin that occur during firing of the missile. In bomb fuzes, activation of the power supply requires that the windmill be moving through air with a speed of at least 100 to 200 mi/hr. While the bomb is in the airplane, an arming wire attached to the airplane prevents the windmill from turning. This wire is withdrawn when the bomb is released, allowing the windmill to turn. In rocket and mortar shell fuzes, the generator cannot turn until the missile is fired. Setback frees the generator shaft but, in addition, a predetermined minimum number of turns of the shaft, under acceleration, is necessary to free it permanently so that the generator will continue to turn after setback. Thus, shocks of short duration, such as may be experienced when the fuze is dropped on a hard surface, will not free the generator. Such an arming system can be designed so closely to a particular acceleration-time characteristic that it will operate only on one selected type of projectile, if desired.

The second stage of arming—completion of the detonator circuit—consists of either unshorting the detonator or connecting it in the circuit, and in most cases (the bomb fuzes are an exception) of charging the detonator-firing capacitor through a time-delay circuit. The actual process of firing the detonator consists of discharging a capacitor through the thyratron and detonator. By this method, considerable current is passed through the detonator quickly, resulting in minimum delay (less than 1 msec) between the triggering of the thyratron and the firing of the detonator. If, however, the thyratron is triggered before the capacitor is charged above a certain minimum voltage, the energy of discharge will be inadequate to fire the detonator and the fuze will be inoperative. By interposing a high resistance between the power supply and the capacitor, completion of the second stage of arming can be delayed as long as desired. This delay is chosen to allow the projectile to be well away from the firing point before arming occurs.

In shell fuzes, the detonator is shorted before arming by a small mercury switch. The spin of the projectile displaces the mercury, unshorting the detonator. In nonrotating fuzes, the detonator is initially disconnected from the circuit. It is rotated into contact with the circuit by means of a low-speed drive shaft, connected through speed-reducing gears to the generator shaft. A certain angular rotation of the low-speed shaft is necessary to connect the detonator. This rotation corresponds to a predetermined number of turns of the windmill of the generator which, in turn, represents a certain air travel of the missile. Thus, a bomb must travel a minimum predetermined distance after release before arming occurs.

The third arming stage—removal of the mechanical barrier—is effected in rotating fuzes by spin and in nonrotating fuzes by the same device that connects the detonator to the circuit. In bomb fuzes, the second and third arming stages occur at essentially the same time, but in fuzes with long delays in the charging of the detonator-firing capacitor, the second stage may be completed after the third stage.

Proximity fuzes are extremely safe, safer than ordinary contact or time fuzes in normal handling, that is, in transportation, storage and ammunition loading. This is because the explosive action is initiated by an electric detonator that obtains its power from the fuze power supply. As already explained, these power supplies are inert until properly activated, as by firing in a gun or releasing from an airplane.

### Development and Production of Proximity Fuzes

It is apparent from the preceding description of the operating principles of radio proximity fuzes that the circuits are really quite simple. Simplicity was, in fact, a cardinal requirement for any design. As ammunition items, proximity fuzes had to be built in tremendous quantities since a fuze is required for each round fired. A complicated design could not have been built in large quantities in time to be of any use in the war; nor could it have been compressed into the very limited volume that was available for the fuze. Compare the fuzes pictured in Figs. 2 and 3 with any miniature radio set and remember that

not only is the fuze a complete transmitter and receiver, but it carries its own electric power supply and the necessary arming and safety features required of ordnance appliances.

Another prime requirement of the fuze was reliability. Since the fuze circuits were activated after firing or release of the missile, no adjustments were possible; dependable operation had to be insured beforehand. Furthermore, testing of, or complicated tinkering with, the fuze at the time of installation in the missile would have rendered the operational use of such fuzes impracticable because of the large scale on which they were to be used.

Development of the fuzes was complicated by the fact that the only completely reliable test was the proof test—that is, use in the missile for which the fuze was designed, under simulated combat conditions. Since such testing usually destroyed, and certainly damaged, the fuze, causes of malfunction often could not be determined by subsequent examination. Thus, good or bad features of design had to be determined by statistical inference, and this required large quantities of fuzes. For this purpose, small production lines had to be put in operation before development was complete. Such lines also served to determine the adaptability of a design to mass production technics, an essential feature of any fuze model.

These factors, coupled with the ever-present time factor, meant that the research, development and production, traditionally carried out in series, were telescoped into a parallel, essentially simultaneous, activity. The fact that the job was done, and millions of reliable fuzes were built, is due not only to the reduction of an intricate problem to fairly simple terms but also to the complete and wholehearted cooperation between the military, the scientists and engineers, and American industry.

The solution of the production problem was probably the most outstanding feature of the proximity fuze program, and it was the nature of that problem which made the proximity fuze project distinctive from other noteworthy wartime developments, such as radar and the atomic bomb. Literally, millions of these electronic devices were produced—devices that could be adequately tested only by destroying them.

Numerous organizations participated in the various phases of the development, production and testing. The relationships among the various contributing groups is most easily shown in chart form, and Fig. 1 is included for this purpose. Two thousand industrial companies participated in the production, which yielded approximately 20 million shell fuzes and 2 million bomb and rocket fuzes.

In the shell fuze program, the most difficult production problem was that of the vacuum tubes and the reserve batteries. Some models of shell fuzes required five tubes, and the minimum number used in any fuze was three. Consequently, these critical items had to be built in appreciably larger quantities than the fuzes. A production rate of 100,000 tubes a day was maintained for long periods, and, near the end of the war, a maximum rate of 400,000 tubes a day was attained. Each of the 44 components of the vacuum tubes had to be precisely fabricated and accurately assembled so that the entire assembly would withstand the terrific impact of gun fire and operate reliably while in flight. Proper operation of each tube, and consequently the fuze, depended on maintaining close tolerance limits for each of the almost microscopic components. Each tube was subjected to inspection and testing at six stages of its manufacture, and upon completion was required to pass a centrifuge test at 20,000 g. Occasionally, several days' entire production had to be discarded because of apparently minor variations in some component or assembly technic. A similar problem was present in the reserve battery for the shell fuze. The numerous individual cell cavities and electrodes of the battery had to be accurately and carefully located so that each cell would receive the proper amount of electrolyte when the gun was fired, thus insuring a proper and steady voltage to operate the fuze.

In the production of fuzes for bombs and rockets, the vacuum tubes and power supplies were also critical items. Microphonics, induced by the air turbulence associated with a high-speed missile in flight, made stability of the vacuum tubes important. The design and production requirements for a microphonically stable vacuum tube were not too dissimilar from the requirements for ruggedness (ability to withstand

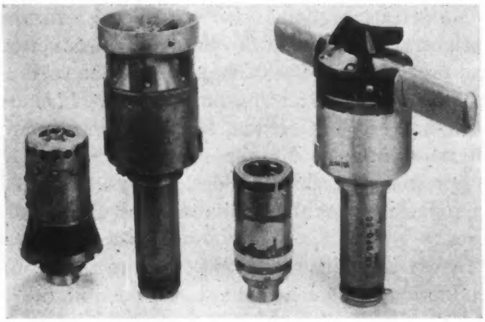


FIG. 10. Representative fuzes for nonrotating missiles. These are, from left to right, rocket fuze, ring-type bomb fuze (longitudinal excitation), 81-mm mortar shell fuze, and bar-type bomb fuze (transverse excitation).

high acceleration). Also, since the generator power supply was a potential source of additional vibration, it had to be accurately dimensioned and carefully balanced.

The full scope of the production program can be appreciated only by analyzing the components of a fuze. In a typical shell fuze, the radio section alone had 40 different components, consisting of more than 300 parts. Most of the components were particularly designed for proximity fuzes, requiring the development of special tools for their fabrication in large quantities. These components were generally made by subcontractors. Final assembly of the fuze was confined to five major plants.

There was similar complexity in bomb fuze production. A typical generator-powered bomb fuze had more than 250 separate and different parts, of which less than 50 were standard items. Final assembly of the fuze was simplified by reducing the design to about 10 subassemblies, of which some were made by the prime contractor and others by subcontractors. Approximately 700 operations were required to build the fuze and its subassemblies.

A further complexity was imposed on production by the number of different models required. Different missiles, or different military applications, frequently required specific fuze designs. For bombs, rockets and trench mortar shells, some 30 different designs were developed, 9 of which were produced in appreciable quantities and an additional 6 of which were in various stages of large-scale production when the war ended. In Fig. 10 are shown representative fuzes for bombs, rockets and trench mortar shells. Diversity was even greater in the shell-fuze program (see Fig. 11); one company alone produced 43 full-scale production models.

The multiplicity of models was necessitated to some extent because rapid development, the urgent time factor, and the concurrent nature of production and development frequently made early fuze models definitely obsolete. For example, the first proximity fuzes were of the photoelectric type rather than radio, and operated when the shadow of the target entered the sensitivity zone of the fuze. Photoelectric fuzes had many limitations (they were primarily restricted to daytime use) but were developed first because the problems involved appeared more amenable to rapid solution than those posed by radio proximity fuzes. However, when production of radio fuzes was realized, photoelectric fuzes were relegated to second place. Similarly, the first radio proximity fuzes were powered with dry batteries because that approach was more expedient. But when reserve batteries and generator power supplies proved practicable for fuzes, the designs with dry batteries became obsolete.

To make even a small change in a device that is being produced in large quantities may result in considerable delay. Tools have to be modified or, in extreme cases, new tools have to be designed and built; new components have to be ordered and stocked, drawings issued, and fresh instructions given to the assembly lines. Thus, with production running concurrently with development, it was essential to have full and complete collaboration of all participating groups in order to get the final models of fuzes to the armed services in time to be of value. Production companies had to make appreciable departures from established procedures in order to secure more flexible operation; laboratory men had to familiarize themselves with the limitations of production and to realize the necessity of providing engineering releases before a development was "perfected"; military men had to make difficult decisions coordinating development possibilities, production limitations, and tactical requirements.

## Testing of Fuzes

*Quality control.*—The production of large numbers of high-quality proximity fuzes required close control and inspection of the basic components, the assembly procedure and the final units. The quality control of ammunition production is usually handled by the Army or Navy,

but the unusual nature of proximity fuzes as ammunition items made it necessary for the developmental laboratories to participate actively in this phase of the work, including the preparation of performance specifications and the setting up of inspection procedures. In addition, laboratory control testing of the final product was carried out in the Applied Physics Laboratory for shell fuzes and at the National Bureau of Standards for the bomb and rocket fuzes.

Representative samples from each production lot or each day's run were sent promptly to these laboratories for critical examination to determine whether the production procedure was satisfactory or in need of some alteration. Although laboratory tests could not be substituted completely for field tests, it was possible to devise laboratory methods of predicting many aspects of field performance. This possibility led to considerable saving in time, manpower and ammunition. Included in the laboratory tests were jolt and vibration tests to simulate rough handling of the fuzes, and exposure tests to simulate performance under extreme climatic conditions, such as excessive humidity and very high and very low temperatures.

Quality-control tests were also conducted on representative production of critical components, particularly the tubes and the power supplies, prior to the release of these items for building into fuzes. Such procedure increased the probability of securing acceptable fuzes by reducing the number of variables that determined the merit of the final assembly.

Throughout the development of quality-control procedures, there was a continual need for compromise between production and allowable tolerance limits. Too tight control on tolerances reduced production to the vanishing point; too loose control brought production up but yielded unacceptable fuzes. One extreme was as bad as the other, and proper compromise required much testing, both in the laboratory and in the field, and numerous trial runs on the pilot production lines. As the development progressed, it was possible to specify designs and tolerance limits that allowed both increased production and highly acceptable quality.

The solution of such problems represents one of the most important technical achievements of the proximity fuze program, and credit for the solution belongs jointly to the military, to industry, and to the laboratories. A full appreciation of such problems makes it evident why the Army and Navy still consider proximity fuzes as secret weapons, even though their general operating principles have been quite fully disclosed. The important quantitative data, obtained only by the expenditure of considerable time, energy and ingenuity, are still military secrets.

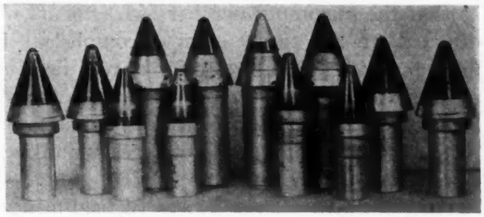


FIG. 11. Representative array of radio proximity fuzes developed for use on antiaircraft and artillery shells. [Applied Physics Laboratory photograph.]

*Field testing.*—As indicated earlier in this paper, the final answer as to the suitability of a proximity fuze design was provided by the field test. Field testing usually consisted of firing the fuze, assembled in a shell, from a gun, or of dropping it, assembled in a bomb, from an airplane and observing the operation of the fuze. The importance of the field test was appreciated very early in the program. A few weeks after development work was initiated in 1940 by Section T, laboratory models of photoelectric bomb fuzes were tested at Dahlgren, Virginia, by dropping them from a bomber. Every stage of the development program was checked with carefully planned field tests. Also, samples of each production lot had to perform properly in the field before the lot was turned over to the services for distribution to the battle areas.

Most of the production testing was done at the Army Proving Ground at Aberdeen, Maryland, and at the Naval Proving Ground at Dahlgren; but various other established proving grounds, such as Jefferson Field, Southwestern, Fort Sill, Fort Bragg and Eglin Field, participated in many phases of the testing and evaluation. In addition, seven proving grounds were set up solely for the testing of VT fuzes: Stump Neck, Maryland; Newtown Neck, Maryland; Fort Miles, Delaware; Fort Fisher, North Carolina; Blossom Point, Maryland; Clinton, Iowa; and the New Mexico proving ground at Albuquerque.

A number of special design problems had to be worked out by field test. Proper burst control design—that is, control of the position of function

so that the maximum number of fragments was directed at the target—was accomplished by firing models of tentative design at real airplanes or simulated airplane targets. Examples of target arrangements used for this purpose are shown in Figs. 8 and 9. Another type of experimental field test, made at the special proving ground at Newtown Neck, Maryland, was vertical firing of shell fuzes for recovery. Most of the important design problems for rugged tubes and reserve batteries were solved at this range. One interesting type of observation which became standard procedure on most field tests was that of listening to the radio transmitter of the fuze while the missile was in flight. Valuable design data were obtained by picking up the radio signal and either listening to the modulations with a loud speaker or recording them on film or disk for later examination.

### Indoctrination

The proper use of proximity fuzes in combat required some training in handling but, more important, required a full appreciation of the possibilities as well as the limitations of use of the fuzes. Because of the secrecy surrounding the development, ordnance men in the combat areas were completely unaware of even the existence of the fuzes until committed to battle. Accordingly, special plans were laid to insure proper use of the fuzes. The plan of the Army Ordnance Department is noteworthy. Almost a year before the first artillery fuzes were used in Europe, carefully selected officers and enlisted men were stationed at the Applied Physics Laboratory, where they worked along with the civilian scientists in learning the ins and outs of proximity fuzes. A smaller group was stationed at the Bureau of Standards for bomb and rocket-fuze training. These men formed the VT Fuze Detachment. The Detachment was divided into teams, and one team was sent to each army in the various theaters of operation. These teams visited combat organizations to give special instructions and then remained in the field to observe combat performance and to serve as consultants and trouble shooters. This indoctrination procedure drew high praise from the combat troops, since effective operational use of the fuzes was accomplished simultaneously on many fronts with a minimum of delay and experimentation.

Other branches of the Service adopted somewhat different procedures, but were aware of the need of some special preparation to make efficient use of the new weapons. The Navy had technical officers at the Applied Physics Laboratory throughout the entire program and, in addition, commissioned some civilian scientists to assist in initial combat operations. A few months before the first bomb fuzes were used, the Army Ordnance Department organized a special course for air ordnance personnel, and the Army Air Corps summoned representatives from each of the combat air forces. These men were given a short, intensive training at the Bureau and at Aberdeen and returned to their organizations to train their personnel in combat use of the fuzes.

Supplementing the indoctrination program, OSRD sent representatives from the developmental laboratories to serve as consultants and observers in the operational theaters.

*Science, by itself, provides no panacea for individual, social and economic ills. It can be effective in the national welfare only as a member of a team, whether the conditions be peace or war. But without scientific progress, no amount of achievement in other directions can insure our health, prosperity and security as a nation in the modern world.*—VANNEVAR BUSH.

# Diffraction Patterns of Microwave Paraboloid Antennas

GORDON FERRIE HULL, JR.*
*Dartmouth College, Hanover, New Hampshire*

THE diffraction patterns of microwave paraboloid antennas resemble in many ways those produced by the diffraction of light through a circular opening. To see how microwave diffraction can be approached from the point of view of diffraction of light, let us first consider the function of the paraboloid antenna. A source of microwave energy, such as the open end of a wave guide, is placed at the focus of a paraboloid of revolution. The source acts as a point source in so far as it emits a spherical wave. The paraboloid serves to convert the spherical wave into a plane wave, which is then propagated along the axis of the paraboloid. This plane wave is then bounded by the edge of the paraboloid in the same way that a plane wave of light striking a circular opening in an opaque screen is bounded by the edge of the opening.

Let us confine our attention to the plane wave at the aperture of the paraboloid. We will assume that the energy in the plane wave is distributed in some manner over the aperture, and that outside the aperture the energy is zero. Let us take for the amplitude distribution function

$$W(\rho, \theta) > 0 \quad \text{for} \quad \rho \leqq \rho_0,$$

and

$$W(\rho, \theta) = 0 \quad \text{for} \quad \rho > \rho_0,$$

where $\rho$ and $\theta$ are polar coordinates in the plane of the paraboloid aperture. Next we apply Huygens' principle to the plane wave by dividing it into a great many spherical wavelets all of the same phase but different amplitudes. We are interested in the intensity, or rather the amplitude, at some point $P_0$ in space a great distance $R$ from the paraboloid antenna. To find this, we obtain the amplitude of one wavelet at $P_0$ and then add up the amplitudes of all the wavelets. It is the interference of these wavelets that produces the paraboloid antenna pattern. The amplitude $I$ at $P_0$ can be expressed by the equation

$$I = C\int_0^{2\pi} d\theta \int_0^{\rho_0} W(\rho, \theta) e^{-2\pi i R/\lambda} \rho d\rho. \qquad (1)$$

The problem is now essentially one of a plane wave of light diffracted by a circular opening, the diameter of which is that of the paraboloid aperture. The derivation of Eq. (1) for the case of $W(\rho, \theta) = 1$ can be found in any standard textbook on optics[1] or theoretical physics.[2] The meanings of $\rho$, $\rho_0$, $\theta$ and $R$ are indicated in Fig. 1, and $\lambda$ is the wavelength. If the amplitude distribution over the paraboloid aperture has circular symmetry, then $\theta$ in $W(\rho, \theta)$ can be omitted. Later it will be shown that the amplitude distribution depends upon the focal length $f$ of the paraboloid, as well as on $\rho$. Hence instead of $W(\rho, \theta)$ we write $W(\rho, f)$. Furthermore, $e^{-2\pi i R/\lambda}$ can be written $e^{ik\rho \cos\theta}$, where $k \equiv (2\pi/\lambda) \sin\varphi$; Eq. (1) then becomes[2]

$$I = C\int_0^{2\pi} d\theta \int_0^{\rho_0} W(\rho, f) e^{ik\rho \cos\theta} \rho d\rho. \qquad (2)$$

The evaluation of the integral (2) will yield the paraboloid antenna pattern. To obtain the gain of the antenna with respect to a point source of unit intensity, we have to evaluate the constant $C$. It can be shown that the total power received from the antenna on the surface of a sphere of radius $R$ is[1]

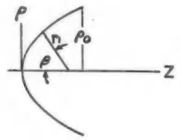
$$C^2 R^2 \lambda^2 \int_0^{2\pi} d\theta \int_0^{\rho_0} |W|^2 \rho d\rho.$$

We will set this equal to the total power $4\pi R^2$ that would be received on the surface of the sphere from a nondirective point source of unit intensity placed at the center of the sphere. Hence

$$C = \frac{(4\pi)^{\frac{1}{2}}}{\lambda} \left[ \int_0^{2\pi} d\theta \int_0^{\rho_0} |W|^2 \rho d\rho \right]^{-\frac{1}{2}}. \qquad (3)$$

* The initial part of this research was carried out while the author was a Member of the Technical Staff of the Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

[1] M. Born, *Optik* (Springer, Berlin, 1933), pp. 141–218.
[2] J. C. Slater and N. H. Frank, *Introduction to theoretical physics* (McGraw-Hill, 1933), pp. 302–328.



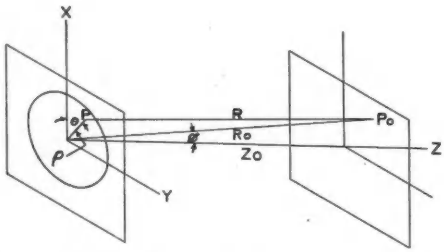




FIG. 1. Polar coordinate system used in the mathematical formulation of diffraction patterns of microwave paraboloid antennas. In the upper diagram, the polar coordinates of the paraboloid are shown, and in the lower diagram the paraboloid aperture is shown as an opening in an opaque screen in the $X$, $Y$ plane. The diffraction effect is obtained at a large distance $R_0$ from the origin of the $X$, $Y$ plane.

When this value of $C$ is substituted in Eq. (2), the gain $G$ of the antenna is given by

$$G=|I_{\max}|^2.$$

This method of calculating diffraction patterns of microwave paraboloid antennas is known as the Kirchhoff-Huygens method. The results obtained from it agree very well with experiment in the case of diffraction of light. Nevertheless, it has been shown by several authors[3] that this method contains fundamental analytic errors. Among these is the error of the assumption that an electromagnetic field can be represented by a single scalar wave function, and the solutions obtained by this method do not satisfy the Maxwell equations. However, the success of the method in the case of light leads one to assume that success in the case of microwaves will also be obtained, provided the wavelength is small compared to the dimensions of the physical apparatus.

Consequently the Kirchhoff-Huygens method has been used to calculate diffraction patterns of paraboloid antennas, first for the case of uniform energy distribution over the paraboloid aperture, and second for the case of a circular wave guide at the focus of the paraboloid. The results of these calculations are then compared with experiment.

### Uniform Energy Distribution Over Paraboloid Aperture

This is the usual case that is discussed in any standard textbook on optics and that corresponds to diffraction of light through a circular opening. For uniform energy distribution over the paraboloid aperture, $W(\rho, f)=1$ and Eq. (2) integrates into

$$I=C\pi\rho_0{}^2\frac{2J_1(x_0)}{x_0},$$

where $x_0=k\rho_0=(2\pi\rho_0/\lambda)\sin\varphi$, $J_1(x_0)$ is the first-order Bessel function, and, from Eq. (3), $C=\sqrt{(4\pi)}/\lambda\sqrt{(\pi\rho_0{}^2)}$. Hence the intensity at any point on the antenna pattern is given by the equation

$$I^2=4\pi\frac{\pi\rho_0{}^2}{\lambda^2}\left|\frac{2J_1(x_0)}{x_0}\right|^2,$$

and the gain of the paraboloid antenna is $G=4\pi(\pi\rho_0{}^2/\lambda^2)$, since

$$\lim_{x_0\to 0}\left|\frac{2J_1(x_0)}{x_0}\right|=1.$$

Absolute values of $J_1(x_0)/x_0$ in decibels down are plotted as a function of $x_0$ in Fig. 2. The height of the minor lobes is independent of the diameter of the aperture and the wavelength. The position of the minor lobes as well as the widths of the minor and major lobes do depend upon the diameter. In fact, the major-lobe beam width is proportional to $\lambda/\rho_0$, as in the optical case. Furthermore, the antenna pattern is entirely independent of focal length.

Next let us examine in a qualitative manner the way in which the antenna pattern will change when the aperture of the paraboloid is not uniformly illuminated. Suppose that this nonuniform illumination of the aperture is produced by a directive source such as the open end of a wave guide placed at the focus of a paraboloid. A curve representing this energy distribution will be higher at the center than at the edges of the aperture. Let us also assume that the energy distribution has circular symmetry with respect

[3] B. B. Baker and E. T. Copson, *The mathematical theory of Huygens' principle* (Oxford Univ. Press, 1939); J. A. Stratton, *Electromagnetic theory* (McGraw-Hill, 1941), pp. 460–464.

to the axis of the paraboloid. Then the total energy radiated by the paraboloid will be represented by the volume formed by revolving the energy-distribution curve about the axis of the paraboloid. If for this volume we substitute a cylinder of equal volume whose height is equal to the average height of the energy-distribution curve, then the rectangle, which when revolved forms this cylinder, represents an equivalent uniform energy distribution. It is easily seen that the width of this rectangular distribution is narrower than the paraboloid aperture. Hence the pattern obtained from a paraboloid antenna with nonuniform illumination over the aperture will be equivalent to the pattern obtained from a paraboloid antenna of smaller aperture diameter having uniform illumination. Consequently, a paraboloid having nonuniform illumination over the aperture, such as that produced by a wave guide at the focus, will have a broader major lobe and less gain than the same paraboloid with uniform illumination over the aperture. However, since the energy at the edges of the paraboloid is less for nonuniform illumination than for uniform, we would expect the minor lobes to be less for the former than for the latter.

### Circular Wave Guide at Paraboloid Focus

Suppose that a circular wave guide containing a $TE_{11}$ wave[4] is placed with its open end at the focus of a paraboloid in such a way that the open end is directed toward the apex of the paraboloid. We will assume that a spherical wave is emitted from the open end of the guide and that the energy distribution over the wave front is that given by diffraction through the open end of the guide. The last assumption is not quite correct because the diameter of the guide is about equal to a wavelength.[4]

We now have two amplitude distributions to consider; first, the distribution due to diffraction through the end of the wave guide, which we assume to be $J_1(x_0)/x_0$, and second, that produced by the paraboloid. To find the latter distribution, let us assume that a nondirective point source is placed at the focus of a paraboloid. Although the energy distribution is uniform over the entire spherical wave emitted by the point

[4] S. A. Schelkunoff, *Electromagnetic waves* (Van Nostrand, 1943), pp. 322 and 357.

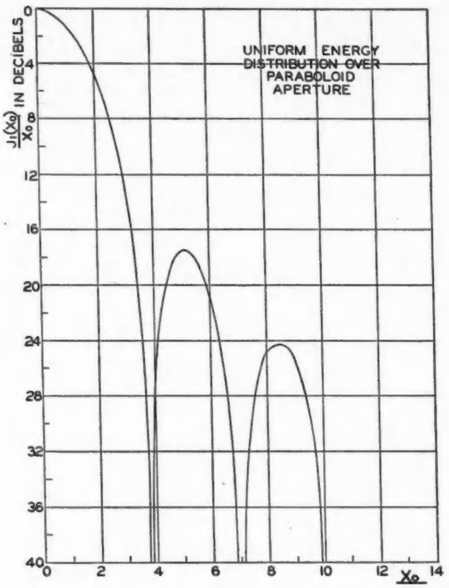



FIG. 2. Diffraction pattern $|J_1(x_0)/x_0|$ of a paraboloid antenna that has uniform energy distribution over its aperture. This corresponds to the optical case of a plane wave of light diffracted by a circular opening in an opaque screen.

source, the plane wave emerging from the paraboloid will not have uniform energy distribution. It can be shown easily that the energy distribution produced by the paraboloid varies with $1/r_1^2$, where $r_1$ is the radius vector from the focus to a point on the paraboloid. Hence the amplitude distribution produced by the paraboloid is

$$\frac{1}{r_1}=\frac{1+\cos\beta}{2f}=\frac{4f}{\rho^2+4f^2}.$$

In Fig. 3, $(1+\cos\beta)/2f$ is plotted as a function of $\beta$, the angle that the radius vector makes with the axis of the paraboloid. As $\rho_0/2f$ increases, the intensity at the edge of the paraboloid decreases. For the case of a paraboloid with the focus in the plane of the aperture, $\rho_0/2f=1$, and the intensity at the edge of the paraboloid is down 6 db from the intensity at the center.

The complete amplitude distribution function now becomes

$$W(\rho, f)=\frac{4f}{\rho^2+4f^2}\frac{J_1(x_0)}{x_0},$$

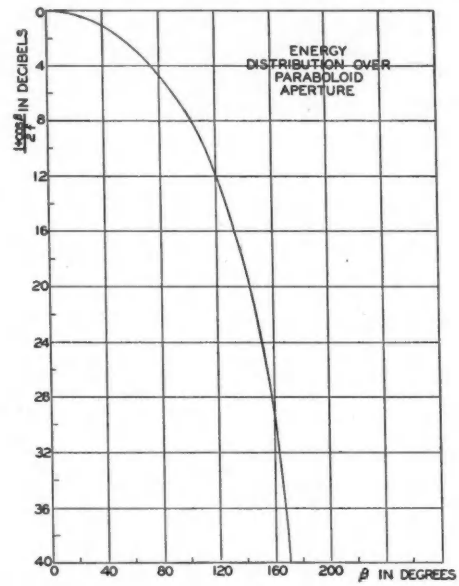



FIG. 3. Energy distribution over a paraboloid aperture when a nondirective point source is placed at the paraboloid focus.

where $x_0 = (2\pi r_0/\lambda)\sin\beta$, and $r_0$ is the internal radius of the wave guide. Since $\sin\beta = \rho/r_1$, as can be seen from Fig. 1, we have

$$W(\rho, f) = \frac{\lambda}{2\pi r_0 \rho} J_1\left(\frac{2\pi r_0}{\lambda}\frac{4f\rho}{\rho^2+4f^2}\right).$$

To find the antenna pattern that results from this distribution we begin with Eq. (2), which integrates immediately with respect to $\theta$, yielding

$$I = C2\pi \int_0^{\rho_0} W(\rho, f) J_0(k\rho)\rho d\rho.$$

Substituting the value of $W(\rho, f)$, we obtain

$$I = C\frac{2f\lambda}{r_0}\int_0^{y_0} J_1\left(\frac{4\pi r_0}{\lambda}\frac{y}{1+y^2}\right) J_0(2fky)dy,$$

where we have set $y \equiv \rho/2f$ and $y_0 \equiv \rho_0/2f$. The integral $I$ cannot be evaluated explicitly unless some kind of approximation can be obtained for

$$J_1\left(\frac{4\pi r_0}{\lambda}\frac{y}{1+y^2}\right).$$

In general an integral of this type,

$$\int_0^{y_0} V(y) J_0(ay)dy,$$

can be evaluated if $V(y)$ can be expressed in a series of the form

$$V(y) = y\left[A + \sum_{n=1}^{n} B_n J_0(C_n y)\right], \text{ for } 0 \leqq y \leqq y_0.$$

The values of $A$, $B_n$ and $C_n$, as well as the number of terms in the series, are governed by the dimensions of the paraboloid. When this series is substituted in the integral, the resulting integrals can be evaluated immediately.[5] The paraboloid dimensions which were chosen are given in Table I.

To evaluate $I$ by means of this approximation method we will choose $r_0 = 0.45$ in. for the internal radius of the wave guide and $\lambda = 3.2$ cm for the

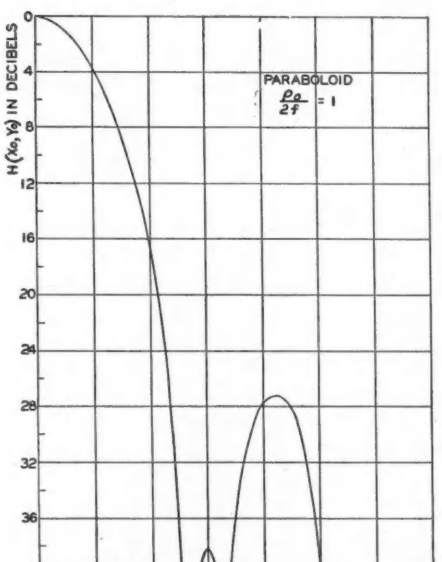



FIG. 4. Theoretical diffraction pattern of a paraboloid antenna when microwave energy is introduced at the paraboloid focus through the open end of a circular wave guide. The shape of the paraboloid is such that the ratio of aperture radius to twice the focal length is unity.

[5] E. T. Whittaker and G. N. Watson, *A course of modern analysis* (Cambridge Univ. Press, 1940), fourth edition, p. 381.

wavelength, which correspond to experimental conditions. Then $4\pi r_0/\lambda = 4.5$, and the amplitude at any point on the antenna diffraction pattern is

$$I = C\frac{2f\lambda}{r_0}\int_0^{y_0} J_1\left(\frac{4.5y}{1+y^2}\right)J_0(2fky)dy,$$

where

$$C = \frac{2\sqrt{2}\pi r_0}{\lambda^2}\left[\int_0^{y_0}\left|J_1\left(\frac{4.5y}{1+y^2}\right)\right|^2\frac{dy}{y}\right]^{-\frac{1}{2}}.$$

For convenience, let us set

$$H(x_0, y_0) \equiv \int_0^{y_0} J_1\left(\frac{4.5y}{1+y^2}\right)J_0(2fky)dy.$$

The function $H(x_0, y_0)$ and the gain $G[=|I_{\max}|^2]$ were calculated for the paraboloid dimensions listed. It was found necessary to use three approximations for $J_1[4.5y/(1+y^2)]$ of two terms each, in order to cover the interval $0 \leqq y \leqq 1$. These are given below, and are accurate to 1 percent:

$$J_1\left(\frac{4.5y}{1+y^2}\right)\begin{cases} = 1.234y + 0.9950yJ_0(5.21y), & 0 \leqq y \leqq 0.43; \\ = 1.065y + 0.8917yJ_0(4.4y), & 0.43 \leqq y \leqq 0.7075; \\ = 0.7430y + 0.6077yJ_0(3.2y), & 0.7075 \leqq y \leqq 1. \end{cases}$$

In Figs. 4 and 5 the absolute value of $H(x_0, y_0)$ is plotted in decibels down as a function of $x_0[=(2\pi\rho_0/\lambda)\sin\varphi]$ for the paraboloid dimensions determined by $\rho_0/2f = 1$ and $\rho_0/2f = 0.7075$. In these figures the wave guide feeding the paraboloid has the ratio of radius to wavelength of 0.358. The antenna diffraction patterns given by $H(x_0, y_0)$ can be applied to any paraboloids at any wavelength provided the foregoing ratios of paraboloid-aperture radius to focal length and wave guide radius to wavelength are maintained,

TABLE I. Paraboloid dimensions.

| $\rho_0$ (in.) | $f$ (in.) | $\rho_0/2f$ |
|---|---|---|
| 9 | 4.5 | 1 |
| | 6.36 | 0.7075 |
| 15 | 7.5 | 1 |
| | 10.6 | 0.7075 |

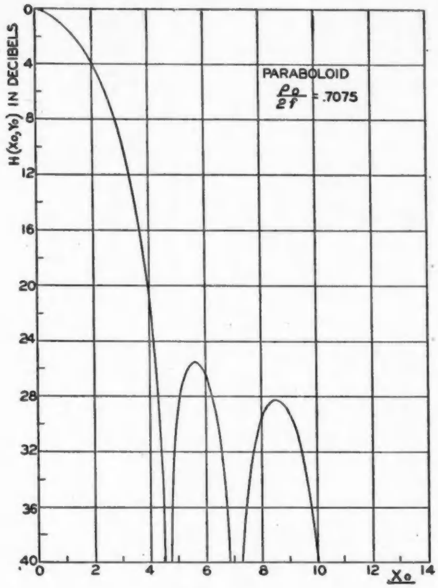



FIG. 5. Theoretical diffraction pattern of a paraboloid antenna when energy is admitted at the paraboloid focus through the open end of a circular wave guide. This pattern is similar to Fig. 4 except that the paraboloid shape is specified by $\rho_0/2f = 0.7075$.

and provided the paraboloid aperture is large compared to a wavelength.

It is interesting to compare the diffraction patterns $H(x_0, y_0)$ in Figs. 4 and 5 with the pattern $J_1(x_0)/x_0$ in Fig. 2, the case of uniform energy distribution over the paraboloid aperture. The major-lobe widths are greater in Figs. 4 and 5 than in Fig. 2, and the minor lobes are lower. The two patterns in Figs. 4 and 5 also differ, and evidently depend upon the value of the ratio $\rho_0/2f$. The larger this ratio is, the wider the major lobe and the lower the minor lobes, which supports the qualitative argument suggested in a previous paragraph.

TABLE II. Paraboloid antenna gain at $\lambda = 3.2$ cm.

| $\rho_0$ (in.) | $f$ (in.) | Gain of paraboloid, aperture uniformly illuminated (db) | Gain of paraboloid, open end of circular wave guide at focus (db) |
|---|---|---|---|
| 9 | 4.5 | 33.1 | 32.3 |
| | 6.36 | 33.1 | 32.6 |
| 15 | 7.5 | 37.5 | 36.7 |
| | 10.6 | 37.5 | 37.0 |

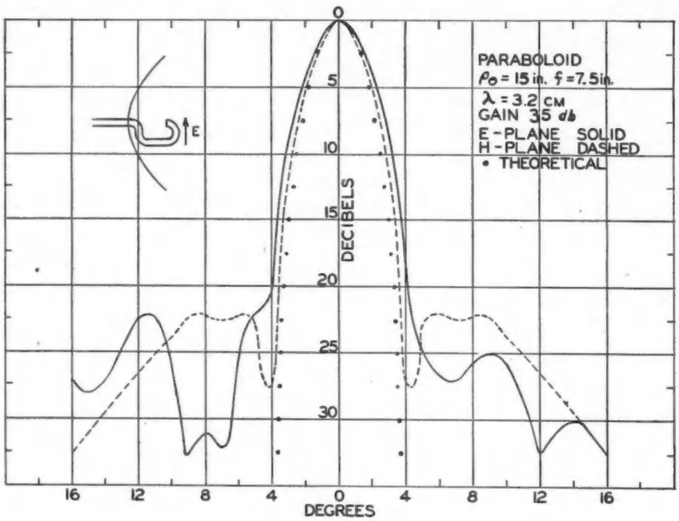



FIG. 6. Experimental diffraction patterns of a paraboloid antenna whose shape is specified by $\rho_0/2f=1$, at a wavelength of 3.2 cm. Microwave energy is introduced at the paraboloid focus by means of the bent wave guide shown in the upper left-hand corner of the figure. The plane of the wave guide bend is in the $E$ plane. The dots represent the theoretical diffraction pattern obtained from Fig. 4.

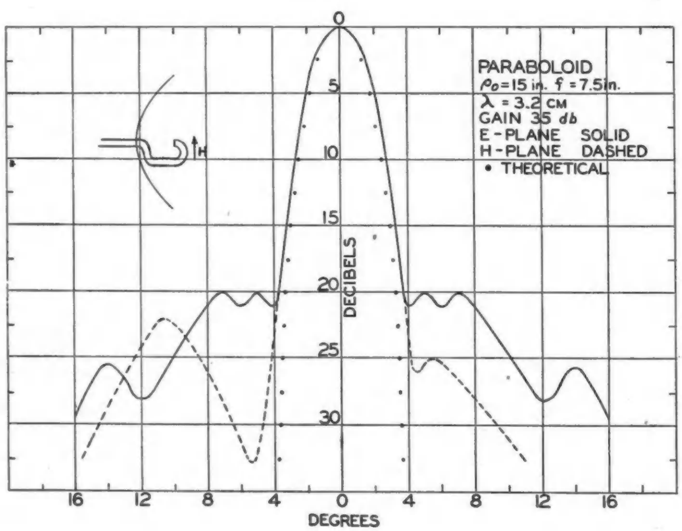



FIG. 7. Experimental diffraction patterns of a paraboloid antenna for $\rho_0/2f=1$ and $\lambda=3.2$ cm. The plane of the wave guide bend is in the $H$ plane. The dots represent the theoretical diffraction pattern obtained from Fig. 4.

The paraboloid antenna gains are listed in Table II, and it will be noted that the paraboloid with uniform illumination over the aperture has larger gain than the paraboloid with a wave guide at the focus. In the latter case the larger $\rho_0/2f$, the less the gain.

### Comparison with Experiment

Paraboloid antenna patterns and gain utilizing two types of wave guide feed systems were measured at a wavelength $\lambda=3.2$ cm, and the results are shown plotted in Figs. 6–9. To make these measurements a microwave transmitter, $\lambda=3.2$ cm, equipped with a 20-db horn,[6] was placed on a platform about 20 ft above the ground and about 50 ft from the paraboloid antenna whose pattern was to be measured. The paraboloid antenna was likewise on a platform 20 ft

[6] G. C. Southworth and A. L. King, Proc. IRE **27**, 95 (1939).

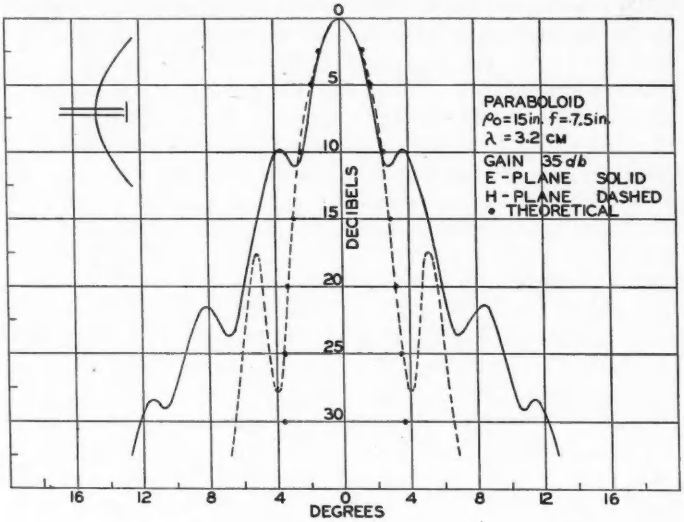



FIG. 8. Experimental diffraction patterns of a paraboloid antenna for $\rho_0=15$ in., $f=7.5$ in. and $\lambda=3.2$ cm. Microwave energy is introduced at the paraboloid focus by a disk reflector in front of the open end of a circular wave guide, as shown in the upper left-hand corner of the figure. The dots represent the theoretical diffraction pattern from Fig. 4.

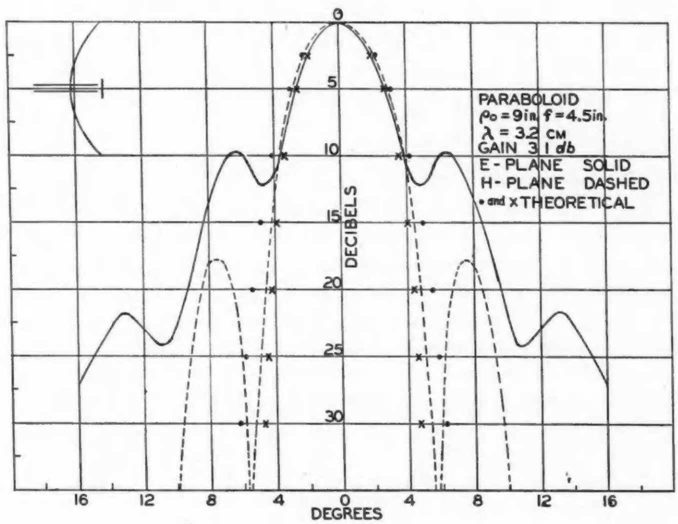



FIG. 9. Experimental diffraction patterns of a paraboloid antenna for $\rho_0=9$ in., $f=4.5$ in. and $\lambda=3.2$ cm, using a disk reflector wave guide feed system. The dots represent the theoretical diffraction pattern of Fig. 4, and the crosses the theoretical pattern of Fig. 2.

above the ground and was connected through a standing-wave detector to a microwave heterodyne receiver. The receiver was equipped with an attenuator having a range of 50 db in variable steps of 1 db. The relative power in decibels received by the paraboloid antenna was measured to an accuracy of $\pm 0.5$ db by varying the attenuator to maintain constant receiver output. The paraboloid antenna pattern was measured by rotating it with respect to the transmitter and recording relative received power as a function of angle. The angular position of the paraboloid antenna was read off a scale marked in half-degree intervals and could be determined to an accuracy of $\pm 0.25°$. Before measurements were made, the antenna was matched to the receiver so that the standing-wave ratio was less than 0.5 db. The paraboloid-antenna gain was measured with reference to a 20-db horn, and is accurate to $\pm 0.5$ db.

Two methods of introducing microwave energy through a wave guide to the focus of the paraboloid were used. The first method consisted of bringing a wave guide 1 in. in diameter through the apex of the paraboloid and bending it so that the open end of the guide was at the focus and pointing toward the paraboloid apex, as indicated in the upper left-hand corners of Figs. 6 and 7. The wave guide bends used in this case had a radius of curvature of 3 in. In Fig. 6 the two experimental antenna patterns for a paraboloid with $\rho_0/2f=1$, one in the plane of the electric vector ($E$ plane), and the other in the plane of the magnetic vector ($H$ plane), are plotted in decibels as a function of angle. For this case the plane of the wave guide bends of the feed system was in the $E$ plane. A similar set of patterns for the case of the plane of the wave guide bends in the $H$ plane are shown in Fig. 7. Comparison of Figs. 6 and 7 shows that the patterns are not identical. The presence of the wave guide feed system produces a pronounced asymmetry in the minor lobes of the pattern in whose plane the wave guide bends lie. The major lobes in Figs. 6 and 7 are also slightly different but exhibit no asymmetry above the 20-db point.

The theoretical diffraction pattern of Fig. 4 is indicated in Figs. 6 and 7 by means of dots. It is seen that the experimental major-lobe width is always greater than the theoretical. Inspection of Fig. 4 also shows that the theoretical minor lobes do not agree either in magnitude or position with the experimental minor lobes in Figs. 6 and 7. This latter disagreement, however, might be attributed to the presence of the wave guide feed system. Also, the experimental gain of 35 db is less than the theoretical gain of 36.7 db.

The second method of introducing microwave energy through a wave guide to the focus of a paraboloid, consisted of bringing a wave guide 1 in. in diameter through the paraboloid apex and reflecting the energy back toward the apex by means of a plane metal disk, as indicated in the upper left-hand corners of Figs. 8 and 9. The disk used was $2\frac{7}{8}$ in. in diameter and was supported $\frac{1}{2}$ in. away from the open end of the wave guide by means of a thin-walled hemispherical polystyrene cup attached to the guide. The disk was located very nearly at the paraboloid focus. In Fig. 8 the $E$ and $H$ plane experimental patterns are plotted for a paraboloid with $\rho_0=15$ in. and $\rho_0/2f=1$. The theoretical pattern of Fig. 4 is shown by means of dots in Fig. 8. In this case the theoretical pattern is in good agreement with the experimental $H$ plane pattern. The theoretical and experimental minor lobes, however, do not agree, as can be seen by inspection of Fig. 4, and the experimental gain is less than the theoretical. In Fig. 9 experimental patterns are plotted for a paraboloid of smaller aperture radius, $\rho_0=9$ in. but with the same ratio $\rho_0/2f=1$ as in Fig. 8. The experimental patterns of Figs. 8 and 9 are similar, but the theoretical dots in Fig. 9 which were obtained from Fig. 4 do not agree with the $H$ plane pattern as they do in Fig. 8. The crosses shown in Fig. 9 were obtained from Fig. 2 and represent the theoretical pattern for the case of uniform energy distribution over the paraboloid aperture. One would expect that, if the theoretical pattern (Fig. 4) agrees with the experimental $H$ plane pattern for a paraboloid of aperture diameter 30 in., it would also agree with a paraboloid of 18-in. diameter, provided the ratio $\rho_0/2f=1$ is maintained, and provided the aperture diameter is large compared to a wavelength. For the smaller paraboloid, a diameter of 18 in. is equivalent to $14\lambda$, which is sufficiently large compared to $\lambda$.

The comparison of the experimental with the theoretical paraboloid-antenna diffraction patterns indicates that the Kirchhoff-Huygens method, in which a single scalar wave function has been used, is incapable of giving quantitative results in agreement with experiment. Occasionally agreement may be achieved, as in Fig. 8, but this case would appear to be fortuitous. Qualitatively, however, the principles of physical optics, as embodied in the Kirchhoff-Huygens method, can be used as a guide in experiment.

*The more we do, the more we can do; the more busy we are, the more leisure we have.*—William Hazlitt.

# Voltage Wave Along a Lossless Line in the General Case

P. LeCorbeiller
*Cruft Laboratory, Harvard University, Cambridge 38, Massachusetts*

**1.** DIAGRAMS showing the distribution of instantaneous voltage (or current) along a lossless line at successive moments of time are well known for the two simpler cases: (*a*) line terminated in its characteristic resistance; (*b*) line open-circuited (or short-circuited). In case (*a*) one has a progressive wave, moving from generator to receiver with the velocity of light $c$; in case (*b*) one has standing waves of amplitude zero at the nodes.

In the general case, when the line is terminated by an impedance $Z_R$ which is neither $R_c$, 0 nor $\infty$, it is usual to plot the rms value of voltage (or current) along the line, thus obtaining an undulating curve (not a sinusoid), the maximum and minimum ordinates of which are in a ratio called the *standing-wave ratio* (dashed lines, Fig. 1).

We have never seen a diagram of successive voltage waves at different moments of time. Such a diagram is represented in Fig. 1. It is seen that at any instant the distribution of instantaneous voltage along the line is given by a sinusoid of wavelength $\lambda = c/f$, this sinusoid progressing from left to right (from generator to receiver) while its amplitude varies from $E_{\max}$ to $E_{\min} = E_{\max}/3$ (taking 3 for the value of the standing-wave ratio).

It appears from this diagram that the wave advances slowly when its amplitude is large and rapidly when it is small. It looks as if the wave, progressing from left to right and constrained between two symmetrical guides (the dashed lines) advanced leisurely when their interval was wide, and slipped rapidly by when it was narrow.

The diagram is more striking if it is continued down to $t = \frac{1}{2}T$. We have limited it here to $\frac{1}{4}T$ in order to save space.

**2.** The calculation of this phenomenon is straightforward. The following formula,[1] valid for all cases, gives the rms complex voltage $E$ at a distance $d$ to the *left* of the load:

$$E = E_R \cos\beta d + jI_R R_c \sin\beta d, \tag{1}$$

$E_R$ and $I_R$ being the complex voltage and current at the load, and $\beta$ denoting $2\pi/\lambda$.

In the general case when the load is an impedance $Z_R$, the distribution of voltage and current is the same as on a line terminated by a pure resistance $R_R = \rho R_c$, where $\rho$ is the standing-wave ratio.[2] We then have $E_{\max} = E_R = \rho R_c I_R$, and Eq. (1) becomes, taking for example $\rho = 3$,

$$E = E_{\max}(\cos\beta d + j\cdot\tfrac{1}{3}\sin\beta d), \tag{2}$$

$d$ being now the distance to the *left* of a voltage antinode, where the voltage is $E_{\max}$.

The instantaneous voltage, a function of distance $d$ and of time $t$, is then

$$\begin{aligned} e(d,t) &= \text{real part of } E_{\max}(\cos\beta d + j\cdot\tfrac{1}{3}\sin\beta d)e^{j\omega t} \\ &= E_{\max}(\cos\omega t\cos\beta d - \tfrac{1}{3}\sin\omega t\sin\beta d). \end{aligned} \tag{3}$$

The sinusoids on Fig. 1 represent the function $e(d, t)$ for successive, constant values of $t$. Putting

$$\cos\omega t = A, \quad \tfrac{1}{3}\sin\omega t = B,$$

and

$$\tan\psi = B/A = \tfrac{1}{3}\tan\omega t, \tag{4}$$

Eq. (3) becomes

$$\begin{aligned} e(d,t) &= E_{\max}\surd(\cos^2\omega t + \tfrac{1}{9}\sin^2\omega t)\cos(\beta d + \psi) \\ &= E_{\max}\surd(\cos^2\omega t + \tfrac{1}{9}\sin^2\omega t) \\ &\qquad \times\cos\frac{2\pi}{\lambda}\left(d + \frac{\lambda\psi}{2\pi}\right). \end{aligned} \tag{5}$$

All the sinusoids on Fig. 1 have thus the same wavelength $\lambda$, and they propagate towards the right with a velocity

$$v_1 = \frac{d}{dt}\left(\frac{\lambda\psi}{2\pi}\right). \tag{6}$$

Now, differentiating Eq. (4), we get

$$(1+\tan^2\psi)\frac{d\psi}{dt} = \tfrac{1}{3}(1+\tan^2\omega t)\omega; \tag{7}$$

[1] King, Mimno and Wing, *Transmission lines, antennas and wave guides* (McGraw-Hill, 1945), p. 18.

[2] Reference 1, Eq. (30.8), p. 37.

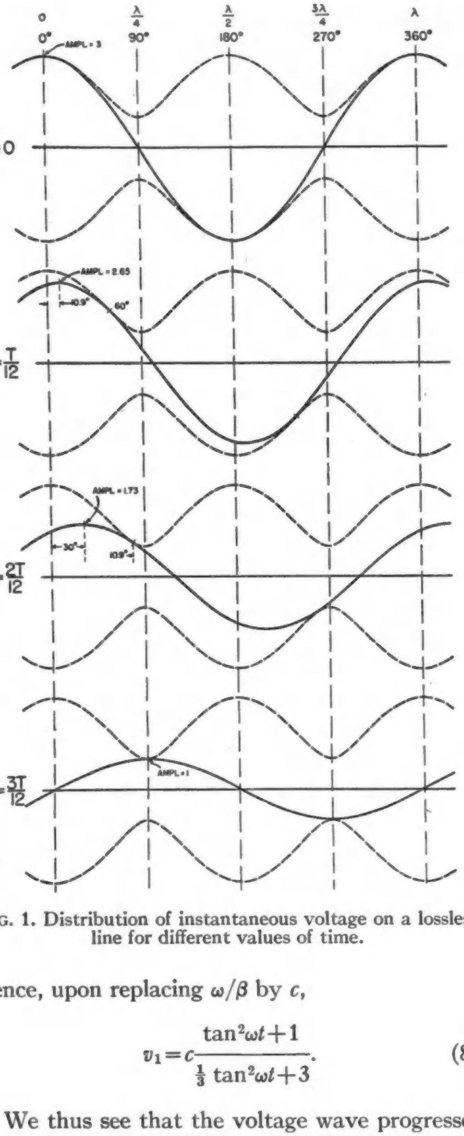



FIG. 1. Distribution of instantaneous voltage on a lossless line for different values of time.

hence, upon replacing $\omega/\beta$ by $c$,

$$v_1 = c\frac{\tan^2\omega t+1}{\frac{1}{3}\tan^2\omega t+3}. \tag{8}$$

We thus see that the voltage wave progresses towards the right with a *small* velocity $\frac{1}{3}c$ at the times $t=0$, $\frac{1}{2}T$, $\cdots$ when its amplitude is large and with a *large* velocity $3c$ at the times $t=\frac{1}{4}T$, $\frac{3}{4}T$, $\cdots$ when its amplitude is small.

**3.** Some time after I had shown him these remarks, Professor R. W. P. King led me to observe that the foregoing result is in apparent contradiction with the usual definition of phase velocity.

The instantaneous voltages at various points on the line are harmonic functions of time of the type $\cos(\omega t+\varphi)$, where $\varphi$ is a function of $d$. The *phase velocity* is the velocity of a fictitious observer who moves along the line so as to "see" always the same value of the phase $(\omega t+\varphi)$.

To find this phase velocity, we now put

$$\cos\beta d = M, \quad \tfrac{1}{3}\sin\beta d = N,$$
$$\tan\varphi = N/M = \tfrac{1}{3}\tan\beta d, \tag{9}$$

and Eq. (3) becomes

$$e(d,t) = E_{\max}\sqrt{(\cos^2\beta d+\tfrac{1}{9}\sin^2\beta d)}\times\cos(\omega t+\varphi). \tag{10}$$

Let $\omega t+\varphi$ be constant, say zero; that is,

$$\omega t+\varphi = 0; \tag{11}$$

then

$$\tan\varphi = -\tan\omega t = \tfrac{1}{3}\tan\beta\delta, \tag{12}$$

calling $\delta$ the distance of the "observer" from the left of a voltage antinode. Differentiating Eq. (11), we obtain

$$\omega+(d\varphi/dt) = 0; \tag{13}$$

and differentiating Eq. (12),

$$(1+\tan^2\varphi)(d\varphi/dt) = \tfrac{1}{3}(1+\tan^2\beta\delta)(d\delta/dt). \tag{14}$$

After a short calculation, taking the phase velocity $v_{ph}$ positive towards the right, we get

$$v_{ph} = -\frac{d\delta}{dt} = c\frac{\frac{1}{3}\tan^2\beta d+3}{\tan^2\beta d+1}. \tag{15}$$

Thus the phase velocity behaves in exactly the opposite way to the velocity $v_1$; it has a *large* value $3c$ at the antinodes $d=0$, $\frac{1}{2}\lambda$, $\cdots$, and a *small* value $\frac{1}{3}c$ at the nodes $d=\frac{1}{4}\lambda$, $\frac{3}{4}\lambda$, $\cdots$.

These variations of the phase velocity can also be seen on Fig. 1, since the point of contact of a sinusoid $t=$const. with its envelope travels towards the right with the velocity $v_{ph}$. To prove this, we remark that to obtain the point of contact we must differentiate Eq. (3) with regard to time, and put $\partial e/\partial t=0$, which gives

$$\partial e/\partial t = E_{\max}(-\sin\omega t\cos\beta d - \tfrac{1}{3}\cos\omega t\sin\beta d) = 0. \tag{16}$$

Thus calling $\delta'$ the distance of the point of contact

to the left of a voltage antinode, we obtain

$$\tan\omega t = -\tfrac{1}{3}\tan\beta\delta',$$

which is precisely Eq. (12).

One actually sees on Fig. 1 that the point of contact traverses a large distance between $t=0$ and $T/12$, and a small distance between $t=2T/12$ and $T/4$, whereas the zero or maximum values traverse a small distance and a large distance, respectively, during the same intervals of time.

**4.** What happens when the load changes continuously from $R_e$ to an open circuit (or a short circuit), causing the standing-wave ratio to increase without limit, provides the mathematical physicist with a neat example of nonuniform continuity, as the reader may enjoy seeing for himself.

# Physics of Rockets: Liquid-Propellant Rockets*

HOWARD S. SEIFERT, MARK M. MILLS AND MARTIN SUMMERFIELD
*Jet Propulsion Laboratory, California Institute of Technology, Pasadena 4, California*

## Liquid-Propellant Chemicals for Rockets

### 22. Advantages of Liquid Propellants

A ROCKET propellant that is liquid rather than solid has two principal advantages. First, most of the propellant may be carried in a light-weight, low-pressure vessel and pumped into the combustion chamber; the latter, though it must withstand high pressures and temperatures, need only be large enough to insure combustion. This results in a significant saving in weight. Second, the flow of liquid may be regulated at will, whereas a solid-propellant motor, with its entire charge in a single high-pressure enclosure, cannot easily be stopped when it is once ignited. A liquid-propellant system has the disadvantage, of course, of being more complicated.

### 23. Important Properties of Liquid Propellants

Because the nature of the liquid used affects the design of a rocket motor, it is appropriate to review the characteristics of typical propellants. Any liquid combination that will release energy at a high but controllable rate immediately after entry into the combustion chamber may be used as a propellant. However, a reaction that proceeds as a detonation is not admissible in the "constant pressure" process we are considering. For best motor performance the propellant should release a maximum of energy, the products of combustion having minimum average molecular weight $M$ and ratio of specific heats $\gamma$. In addition to these basic requirements, an acceptable propellant is hedged about with so many practical restrictions that the search for suitable liquids is one of the largest problems of rocket research, and the curious properties of these liquids account for a large portion of the difficulties of rocket-motor development.

A propellant liquid should have the following properties:

(i) Its *heat of combustion* should be maximum in order to secure maximum chamber temperature.

(ii) The *molecular weight of the combustion products* should be minimum to secure maximum exhaust velocities.

(iii) The liquid should be *stable against shock and temperature changes;* that is, it should not decompose or detonate under mechanical impact or moderately high temperatures.

(iv) Its *rate of reaction* should be high to keep the volume of the combustion chamber small.

(v) The propellant components should *ignite easily*, that is, within a minimum time interval after contact with each other or with the ignitor flame.

TABLE III. Typical liquid propellants.

| Oxidizer | Fuel |
|---|---|
| *Bipropellants* | |
| Liquid oxygen | Ethyl alcohol plus water<br>Ammonia<br>Hydrazine<br>Hydrogen |
| Nitric acid | Aniline<br>Furfural alcohol |
| Hydrogen peroxide | Nitromethane<br>*C*-Stoff (German) |
| *Monopropellants* | |
| Hydrogen peroxide | |
| Nitromethane | |

* The preceding article [*Am. J. Physics* **15**, 1 (1947)] dealt with the principles of rocket propulsion and solid-propellant rockets.

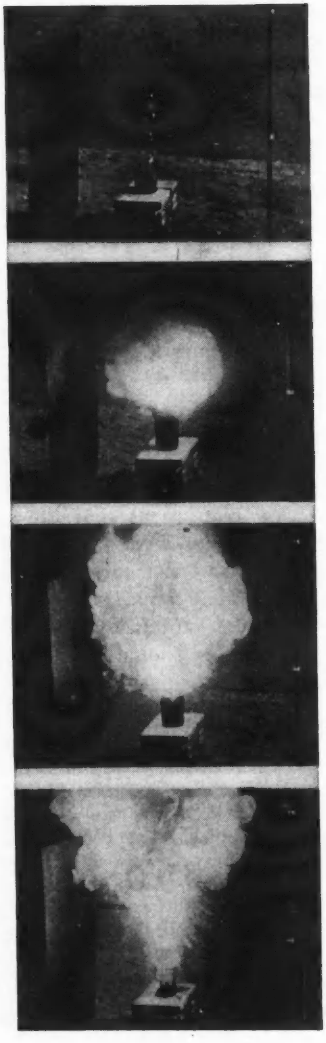


FIG. 18. Consecutive motion picture frames taken $\frac{1}{64}$ sec apart of the spontaneous ignition of red fuming nitric acid and aniline when brought in mutual contact.

(vi) The *density* of the liquid should be high, since such liquids require less tank structure and are easier to pump, and smaller tanks reduce air drag.

(vii) The *vapor pressure* should be low to avoid loss of fluid and the necessity for elaborate thermal insulation, as well as to improve pumping properties.

(viii) The *specific heat and thermal conductivity* should be high if the liquid is to serve as both coolant and propellant.

(ix) The *freezing point* should be low for liquids that are to be used in any geographic location.

The following additional properties are also of practical importance: (x) corrosivity; (xi) toxicity; (xii) inflammability, especially in the vapor phase; (xiii) availability; (xiv) cost.

**24. Tabulation of Typical Characteristics**

A certain rather limited number of liquids have been found that are satisfactory in most of the foregoing respects. None of them is ideal, and the search for still better substances continues. Some propellants consist of a single liquid; these are termed "monopropellants." Those involving two liquids are called "bipropellants." Bipropellants normally consist of an oxidizer and a fuel. The ratio of mass flow rate of oxidizer to that of fuel is called the *mixture ratio r*. The mixture ratio sometimes deviates from the stoichiometric value in order that lower reaction temperature or molecular weight of products may be secured. In Table III are listed ten typical propellants classified by the nature of the oxidizer.

Each propellant has its unique characteristics, some of which are described in the following paragraphs. Quantitative performance data are given later, in Table IV.

*Liquid oxygen–ethyl alcohol.*—This classical combination has the advantage of high specific impulse; moreover, its components are nontoxic, noncorrosive, and nondetonative. However, liquid oxygen has a high vapor pressure at ordinary temperatures, which makes it difficult to store and of little value as a coolant (see SEC. **32**). In the V-2 the Germans used a fuel mixture of 75 percent ethyl alcohol ($C_2H_5OH$) and 25 percent water. This water "ballast" decreases both the reaction temperature and the average molecular weight. Thus the cooling of the motor is rendered easier without materially reducing its performance.

*Liquid oxygen–noncarbonaceous fuel.*—The elimination of carbon from the fuel component tends to reduce the molecular weight of the exhaust gases, since the fuel may then be composed largely of nitrogen and hydrogen. For example, *ammonia* ($NH_3$) is a readily available substance which, calculations indicate, should give high performance. It is toxic and must be confined under pressure in order to remain liquid at ordinary temperatures. A more convenient fuel is *hydrazine* ($N_2H_4$), which is a liquid at room temperature and gives high performance at remarkably low combustion temperatures (see Table IV). Neither of these liquids is a very good coolant—the ammonia because of its low boiling point, and the hydrazine because it decomposes at temperatures in excess of a few hundred degrees Fahrenheit.

*Liquid oxygen–liquid hydrogen.*—Liquid hydrogen gives the highest performance of any available fuel in combination with liquid oxygen. However, its extreme volatility (boiling point approximately −425°F) and low specific gravity (0.07), combined with its relative scarcity, have

prevented its exploitation so far. It is even less suitable as a coolant than liquid oxygen. However, it remains interesting to rocket research workers, representing a sort of ultimate performance goal which they hope to achieve.

*Nitric acid–aniline.*—These components have the virtue of igniting spontaneously upon contact, thus eliminating the necessity for special ignition devices. Figure 18 illustrates the nature of this spontaneous ignition. In order to make ignition more prompt, 6 to 14 percent of nitrogen dioxide ($NO_2$) is dissolved in the nitric acid ($HNO_3$) to make "red fuming" nitric acid, and the water content is kept below a maximum of 2 or 3 percent. The corrosive acid must be handled in stainless steel or aluminum containers, and its vapor is toxic. These disadvantages are partially offset by the high specific gravity (1.55) of the acid and its stability to shock and temperature. Considerable experience in using acid in rocket motors has been accumulated in the United States.[18]

The aniline ($C_6H_5NH_2$) component usually has about 20 percent of furfural alcohol added to it to depress the freezing point. Aniline is toxic but not otherwise difficult to handle. It has a high boiling point and makes an excellent coolant fluid.

*Nitric acid–furfural alcohol.*—This combination is also spontaneously igniting, with rather less ignition lag than in the acid–aniline case. Here the acid need not contain $NO_2$, but may be the so-called "white acid." The furfural alcohol ($C_4H_3O \cdot CH_2OH$) has a low freezing point and is nontoxic. This combination gives a performance equivalent to that of the nitric acid–aniline combination and is somewhat more convenient to pump because the lower vapor pressure of white acid as compared with that of red fuming acid reduces cavitation difficulties.

*Hydrogen peroxide.*—Hydrogen peroxide of 80 to 90 percent concentration is an excellent oxidizer, but is rather unstable when heated and very sensitive to contaminants, especially to the metal oxides, which act as catalysts for the exothermic decomposition $H_2O_2 \rightarrow H_2O + \frac{1}{2}O_2$. Concentrated hydrogen peroxide decomposes with explosive violence at temperatures only a little above the boiling point of water. The freezing point ranges from $-10°$ to 10°F as the concentration increases from 80 to 90 percent. The liquid may be stored for extended periods provided it is pure to within a few parts per million, and is kept in correspondingly clean containers. Pure aluminum containers are best, although nickel, stainless steel and certain vinyl plastics are acceptable.

The controlled decomposition of $H_2O_2$ is usually effected by adding to it a few percent of concentrated calcium permanganate catalyst solution, or by passing it over solid stones containing permanganate or lead compounds. When decomposed in this manner it serves as a low-temperature rocket propellant providing half the specific impulse at about one-fourth the reaction temperature of other standard propellants. It also may serve as a convenient gas generator for turbine-driven pumps.

[18] The original development was carried out at the Jet Propulsion Laboratory, GALCIT. GALCIT is the abbreviation for Guggenheim Aeronautical Laboratory, California Institute of Technology.

*Nitromethane.*—Nitromethane ($CH_3NO_2$), the simplest of the nitroparaffins, contains all of the elements necessary for combustion in the same molecule. It may, therefore, be used as a monopropellant. It must be ignited in the presence of oxygen and will burn smoothly provided a catalyst is added; the combustion pressure is about twice the typical value of 300 lb/in.$^2$, and the available combustion volume is several times[19] the value used with other propellants.

Nitromethane is noncorrosive, nontoxic, insensitive to contamination, and has a low vapor pressure. The fact that only one fluid need be used makes possible a simple tank and valve system. However, it decomposes with explosive rapidity above 550°F and can be made to detonate under mechanical shock. Consequently it can be used as a coolant only with caution.

When nitromethane is used in conjunction with hydrogen peroxide as a bipropellant, it will react at lower pressures (300 lb/in.$^2$) and volumes ($L^* = 100$ in.) than when used alone. Furthermore, it is no longer necessary to add a catalyst to the nitromethane, and ignition can easily be accomplished without a flame or spark, by initiating the peroxide deomposition with a small amount of permanganate catalyst. This combination is a rather convenient propellant because of the nontoxic, noncorrosive nature of its constituents. It should be useful in situations where the components do not need to be preheated excessively.

*C-Stoff.*—In their Me163B rocket-propelled fighter airplane, the Germans used a fuel designated as "*C*-Stoff" which contains 30 percent hydrazine hydrate ($N_2H_4 \cdot H_2O$), 57 percent methyl alcohol and 13 percent water. It ignites spontaneously with $H_2O_2$ by virtue of the hydrazine hydrate content, and is stable enough to be used for cooling the rocket motor.

### 25. Upper Limit of Chemical Propellant Performance

A comparison of the performance and characteristics of the various propellants discussed is given in Table IV, from which it is seen that, despite the great variation in propellants, the difference in performance is not more than 50 percent. It is sometimes hoped that the performance of a rocket can be greatly improved by the discovery of a new propellant. However, thermodynamic calculations, based on the known properties of possible liquid propellants, set a theoretical upper limit on specific impulse which exceeds that of the best now obtainable by only about 80 percent. The limitation is fundamental, and follows from the fact that any propellant consisting of the elements H, C, O and N produces combustion products which dissociate and absorb energy so rapidly above 4000°F that the combustion temperature is limited to values less

[19] For $CH_3NO_2$, $p_c = 550$ lb/in.$^2$ and $L^* = 350$ in. are typical values; see the definition of $L^*$ in Sec. **28**.

TABLE IV. Calculated performance of typical liquid propellants. All values of the exhaust velocity $c$ have been calculated for 300-lb/in.$^2$ combustion pressure, correctly expanded to 1 atmos. Subtract 10 percent to get experimental values of $c$, $c^*$ or $I_{sp}$ from calculated values.

| Propellant | Specific impulse, $I_{sp}$ (sec) | Exhaust velocity, $c$ (ft/sec) | Characteristic velocity, $c^*$ (ft/sec) | Mixture ratio, $r$ | Chamber temp. $T_c$ (°F) |
|---|---|---|---|---|---|
| Liquid oxygen–75% alcohol, 25% water | 239 | 7698 | 5537 | 1.3 | 5079 |
| Liquid oxygen–hydrazine | 246 | 7915 | 5610 | 0.33 | 3632 |
| Liquid oxygen–ammonia | 255 | 8220 | 5840 | 1.4 | 4951 |
| Liquid oxygen–liquid hydrogen | 358 | 11550 | 8345 | 3.0 | 4290 |
| Hydrogen peroxide (87%) | 126 | 4065 | 2940 | — | 1216 |
| Hydrogen peroxide (87%)–nitromethane | 229 | 7386 | 5313 | 0.5 | 4687 |
| Hydrogen peroxide (87%)–"C-Stoff" | 215 | 6921 | 4945 | 2.5 | 3711 |
| Red fuming nitric acid–aniline | 221 | 7091 | 5015 | 3.0 | 5065 |
| White nitric acid–furfural alcohol | 214 | 6885 | 4982 | 1.9 | 4750 |
| Nitromethane | 218 | 7020 | 5020 | — | 3950 |

than 6000°F. Since the limitation is imposed by the nature of the products of combustion, variations in the chemical composition of the propellants cannot produce remarkable increases in specific impulse. The theoretical maximum specific impulse to be expected from liquid oxygen and hydrazine, to mention one little-explored bipropellant, would be about 260 sec. Liquid hydrogen combined with either liquid oxygen or liquid fluorine would give a specific impulse of about 350 sec, which may be considered as the chemical ceiling of performance. Since liquid hydrogen has a specific gravity of only 0.07, it decreases the impulse per unit *volume* of any bipropellant of which it is a component.

The gain of 25 to 50 percent in the specific impulse $I_{sp}$ to be anticipated in the near future is well worth working for. But miraculous chemical propellants that increase $I_{sp}$ by a factor of ten over present values do not appear on the horizon. We may conclude then that the choice of propellant in the future will be decided more by utilization factors than by performance.

With the advent of nuclear energy, this conclusion must be modified. However, if it is still necessary to have a working fluid, and if the heat that can be introduced into this fluid is limited by temperature, it would not be possible to increase the values of $I_{sp}$ by as much as one order of magnitude over current experimental values (see SEC. 46). For example, $H_2$ heated to 4000°K (7200°R) and expanded by the normal adiabatic process yields a specific impulse of 700 sec. If the $H_2$ were dissociated to 2H, the value of $I_{sp}$ might increase to approximately 1000 sec. These predictions could be too pessimistic in that other new methods of obtaining a high-speed jet may be possible. However, the problem of adapting nuclear energy to rocket propulsion is a difficult one and will require years of research and development.

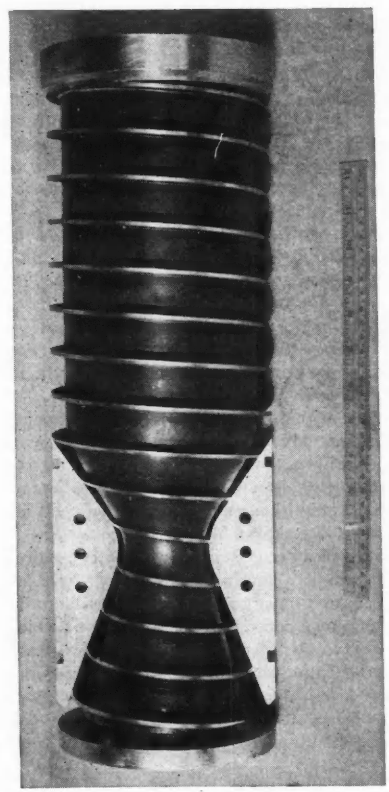


FIG. 19. A 1500-lb thrust rocket motor chamber, showing cooling ducts.

## Fundamentals of Liquid-Propellant Rocket Motors

### 26. Mechanical Design

A typical liquid-propellant, thrust-producing component—which will henceforth be called a *motor*—is illustrated in Figs. 19 and 20, in which the three major parts are indicated. These parts are the propellant injector, the combustion chamber and the exhaust nozzle. The motor shown was designed according to the theory of Part I* for use in a sounding rocket, and can be described by the following data:

| *Performance data (sea level)* | |
|---|---|
| Thrust, $F$ | 1500 lb |
| Combustion pressure, $p_c$ | 300 lb/in.$^2$ |
| Specific impulse, $I_{sp}$ | 193 sec |
| Exhaust velocity, $c$ | 6200 ft/sec |
| Duration of operation | 45 sec |
| Motor weight | 50 lb |
| Chamber temperature, $T_c$ | 5000°F |
| Propellants | acid–aniline |
| Mixture ratio, $r$ | 2.75 |
| Coolant | aniline (fuel) |

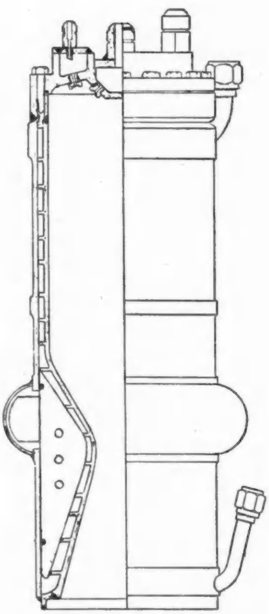


FIG. 20. Section of 1500-lb thrust rocket motor, showing, from top to bottom, propellant injector, combustion chamber and exhaust nozzle.

* *Am. J. Physics* **15**, 1 (1947).

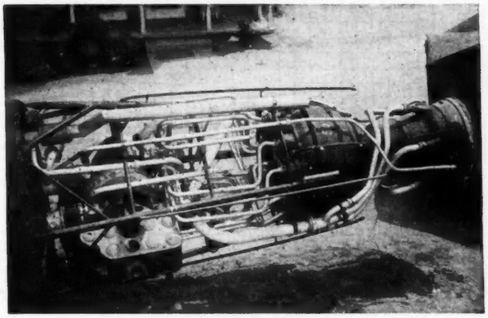


FIG. 21. Exposed aft end of V-2 rocket, showing motor contour and elaborate piping installation.

| *Geometric data* | |
|---|---|
| Throat diameter, $d_t$ | 2.17 in. |
| Exit diameter, $d_e$ | 4.85 in. |
| Over-all length | 22.5 in. |

The motor is fabricated of a type of stainless steel that has relatively high thermal conductivity. It is provided with an expansion joint to allow for the high operating temperatures of the inner shell, which is welded rigidly to the outer shell. The aniline coolant is conducted spirally around the nozzle and combustion chamber before entering the injector manifold (regenerative cooling). The motor interior is chrome plated for resistance to corrosion and erosion.

### 27. Effect of Application on Design

The following requirements are those that most strongly influence the design of the rocket motor intended to fit a specific application: (i) magnitude of thrust; (ii) duration of continuous operation; (iii) operation expendable or repeatable; (iv) operating altitude; (iv) permissible motor weight.

(i) *Magnitude of thrust.*—Motors having less than 100 lb thrust are subject to clogging of the minute injection orifices. As the thrust increases above 1000 lb the ratio of length to diameter of the motor decreases, as is evident, for instance, from a comparison of Figs. 20 and 21. So long as a suitable cooled vessel can be built to withstand a pressure of about 500 lb/in.$^2$ in both tension and compression, there appears to be no fundamental upper limit to the size of a rocket motor. A motor delivering a thrust of $10^6$ lb, for example, would have a throat diameter of about 5 ft. Recent studies indicate that in the larger motors, considerable reduction in relative combustion volume can be made without much loss in the combustion parameter $c^*$. This results in a motor having a "morning-glory" shape, consisting of a tubular section followed by a slightly constricted throat and an expanding cone.

(ii) *Duration of continuous operation.*—A rocket motor of the type shown in Fig. 20 usually reaches thermal

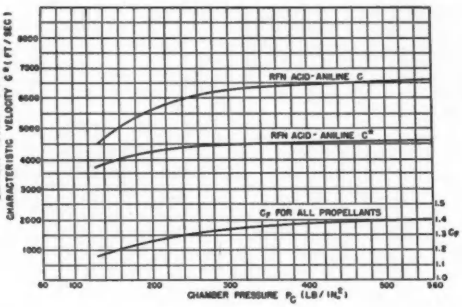



FIG. 22. Variation of propellant characteristic velocity $c^*$ with combustion pressure $p_c$ for nitric acid and aniline. $r=2.75$.

equilibrium in about 30 sec. If the motor is to be used for periods shorter than this it may be built with heavy uncooled walls, thermal capacity being depended upon to keep it at a safe operation temperature. The duration of operation of a cooled motor is limited only by the amount of propellant available, since erosion of the nozzle throat —the vulnerable point at which highest heat transfer occurs—usually does not occur before exhaustion of the propellant. The limit of duration of a rocket-powered airplane, determined by the propellant weight which it can carry per pound of rocket thrust developed, is of the order of 1 hr or less. If there is no assisting lift from the air, as in the case of a wingless missile rising vertically, the operation period is evidently limited by the necessity that the weight of propellant be less than the thrust. It is interesting to note here that a vertical acceleration of $2g$ must be maintained for 10 min in order that a rocket may escape the gravitational field of the earth.

Certain ranges of duration of motor operation are typical of specific applications. They are illustrated in Table V.

(iii) *Repeatable versus expendable operation.*—In expendable service it is often possible to tolerate a certain amount of erosion and build a much lighter motor than would be required for repeated operation, particularly if no human passenger is involved, in which case safety factors may be reduced. If the motor need not be started and stopped at will, conventional valves may be replaced by burst diaphragms, resulting in simplification and reduction in weight.

(iv) *Expected operating altitude.*—As was pointed out in SEC. 9, for each external pressure there exists an optimum nozzle whose ratio of exit to throat area gives maximum thrust. Consequently a rigid nozzle can be designed correctly for only one altitude. Thus a motor for sea-level use might have an expansion ratio $\epsilon$ of 3.5, whereas one for use at 40,000 ft would have a ratio $\epsilon$ of 11.0. For vertical flight a mean altitude is chosen. For motors that are to operate outside the atmosphere, the expansion ratio is made as large as is consistent with structural limitations.

(v) *Permissible motor weight.*—For aircraft propulsion, the motor weight may be only 1 or 2 percent of the total, and rather heavy durable construction may be used. For missiles, the motor may have 3 to 10 percent of the total weight, and it is desirable to have the lightest possible construction. With present technics, motors ranging from 20 to 100 lb thrust per pound of motor weight have been built.

### 28. Characteristic Length $L^*$ and Combustion Time

The purpose of the combustion chamber is to provide sufficient volume to allow time for reasonably complete burning of the propellant before the products of combustion reach a cross section near the throat of the exhaust nozzle. In contrast with the nozzle, it is not possible to predict the optimum dimensions of the combustion chamber. For convenience in manufacture, the chamber is usually made in the form of a cylinder of volume $V_c$ and length $l_c$.

It is found experimentally that the ratio $V_c/f_t$ of volume $V_c$ to throat area $f_t$, which we shall define as the characteristic length $L^*$, must exceed a certain minimum value for adequate combustion of liquid propellants. This value may range from 25 to 600 in., depending on the propellant and its method of injection.

A relation showing that the time of combustion $t_c$ during which the reactants remain in the chamber is proportional to characteristic length $L^*$ can be derived on the basis of the following simplifying assumptions:

(i) The propellant components are completely mixed at the injector end of the cylindrical chamber; and sufficient, although incomplete, combustion takes place so that the fluid leaving the vicinity of the injector may be considered to be a gas.

(ii) The velocity and temperature of the fluid, after leaving the vicinity of the injector, are uniform throughout the combustion chamber, even though, as we know, further chemical reaction is taking place.

The combustion time $t_c$ is

$$t_c=l_c/v_c, \tag{80}$$

where $v_c$ is the velocity of the reactants parallel to the axis of the cylindrical chamber. Now from the equation of continuity [Eq. (29)],

$$v_c=\dot{m}/\rho_c f_c=\dot{m}R_s T_c/f_c p_c, \tag{81}$$

and from Eq. (39), $\dot{m}=\Gamma' f_t p_c/a_c$. Putting this value of $\dot{m}$ in Eq. (81) and the resulting value of $v_e$ in Eq. (80), we obtain, with the help of Eq. (52), which defines $c^*$,

$$t_c=\frac{l_c f_c}{f_t}\cdot\frac{a_c}{\Gamma' R_s T_c}=\frac{\gamma L^*}{\Gamma' a_c}=\frac{\gamma L^*}{(\Gamma')^2 c^*}, \qquad (82)$$

where we have assumed that $l_c f_c$ is equal to $V_c$, the effective chamber volume. From Eq. (82) we see that the time during which the reactants remain in the chamber is directly proportional to the characteristic length $L^*$ and inversely proportional to the characteristic velocity $c^*$. For example, a typical acid–aniline motor with 1000-lb thrust will have $c^*=4500$ ft/sec, $L^*=3.33$ ft and $\gamma=1.25$, giving $t_c=0.0017$ sec. As the size and thrust of the motor increase, the value of $L^*$ that is necessary does not increase correspondingly, so that the volume of the chamber becomes a smaller fraction of the total volume and that of the nozzle a larger fraction. This is apparent if one compares the 1500-lb thrust motor shown in Fig. 19 with the 56,000-lb (V-2) motor of Fig. 21.

### 29. Typical Motor Calculation

Three sets of quantities that must be known accurately in order to design a regeneratively cooled liquid rocket motor are the chamber and nozzle dimensions, the hydraulic and mechanical parameters of the injector, and the hydraulic and thermal parameters of the cooling ducts. Consider first the procedure for finding the basic motor dimensions.

The quantities that may be selected more or less arbitrarily are: thrust, $F$ (lb); combustion pressure, $p_c$ (lb/in.$^2$); external pressure, $p_0$ (lb/in.$^2$); propellant chemicals; mixture ratio, $r$. Data

TABLE V. Range of duration of rocket operation.

| Application | Duration (sec) |
|---|---|
| Artillery rockets | 0.1–1.0 |
| Missile launching | 0.5–5.0 |
| Aircraft take-off | 10–45 |
| Missile propulsion | 30–300 |
| Aircraft propulsion | 300–3600 |

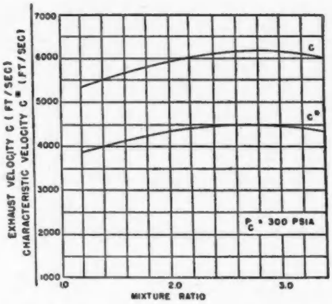



FIG. 23. Variation of propellant characteristic velocity $c^*$ with ratio $r$ of oxidizer to fuel flow rate for nitric acid and aniline.

concerning the performance parameters of the propellant (see Table IV) must be collected empirically beforehand.

From the given conditions, we wish to find values of: throat area, $f_t$ (in.$^2$); exit area, $f_e$ (in.$^2$); chamber volume, $V_c$ (in.$^3$); propellant weight flow rate, $\dot{m}g$ (lb/sec). One proceeds to do this in the following sequence:

(i) Select a value of thrust $F$, combustion pressure $p_c$, external pressure $p_0$; determine the propellant combination and, if a bipropellant the mixture ratio $r$.

(ii) Using empirical data collected from static rocket-motor tests with the chosen propellant at the values of $p_c$ and $r$ just chosen, in conjunction with thermochemical calculations, determine the ratio of specific heats $\gamma$, characteristic length $L^*$ and characteristic velocity $c^*$. The value of $\gamma$ is between 1.2 and 1.3 for the propellants described in Table IV, while $L^*$ often lies between 50 and 100 in. Typical curves of $c^*$ as a function of combustion pressure and mixture ratio are given in Figs. 22 and 23. These curves indicate that $c^*$ varies rather slowly with $p_c$ and $r$.

(iii) Use the ratio of specific heats $\gamma$ and the pressure ratio $p_c/p_0$ to calculate the nozzle coefficient $C_F$ and nozzle expansion ratio $\epsilon$, as in Eqs. (45) and (50). Graphs of $C_F$ and $\epsilon$ are shown in Fig. 7. These theoretical values of $C_F$ must be corrected slightly for friction and divergence of the nozzle.

(iv) Use the basic relation $F=C_F p_c f_t$ [Eq. (48)] to compute the throat area $f_t$, and the known ratio $\epsilon[=f_e/f_t]$ to compute the exit area $f_e$.

(v) Use the definition [Eq. (52)] of characteristic velocity $c^*$ to calculate the total mass flow rate $\dot{m}$; thus, $\dot{m}=p_c f_t/c^*$. The individual mass flow rates of oxidizer, $\dot{m}_0$, and fuel, $\dot{m}_f$, can then be computed from $\dot{m}[=\dot{m}_0+\dot{m}_f]$ and the mixture ratio $r[=\dot{m}_0/\dot{m}_f]$.

(vi) Determine the combustion volume $V_c$ from the relation $V_c=L^* f_t$, using the empirically measured characteristic length $L^*$.

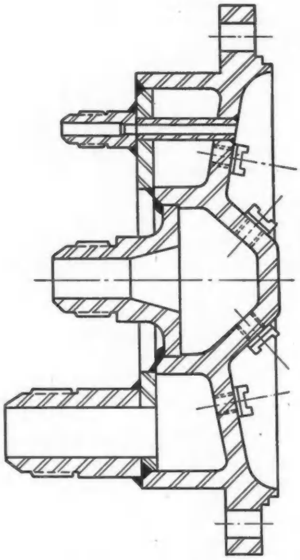


FIG. 24. Section of a typical bipropellant impinging stream injector for a 1500-lb thrust rocket.

(vii) The ratio of combustion-chamber cross-sectional area $f_c$ to throat area $f_t$ is determined from a knowledge of heat transfer and structural strength factors. Heat transfer is increased if $f_c$ is small; stress is increased if $f_c$ is large. A typical compromise value of $f_c/f_t$ is 6. When this ratio is fixed, the chamber length $l_c$ is also determined. Using as an example the 1500-lb thrust motor described in SEC. 26, one arrives by this process at the following quantities, listed in the order in which they must logically be calculated.

| | |
|---|---|
| *Information given:* | |
| Thrust at sea level, $F$ | 1500 lb |
| Combustion pressure, $p_c$ | 300 lb/in.$^2$ |
| Average external pressure, $p_0$ | 8.5 lb/in.$^2$ |
| Propellant | acid–aniline |
| Mixture ration, $r$ | 2.75 |
| *Derived data:* | |
| Ratio of specific heats, $\gamma$ | 1.25 |
| Characteristic velocity, $c^*$ | 4600 ft/sec |
| Characteristic length, $L^*$ | 73.4 in. |
| Corrected thrust coefficient, $C_F$ | 1.35 |
| Expansion ratio, $\epsilon$ | 5.0 |
| Throat area, $f_t$ | 3.7 in.$^2$ |
| Exit area, $f_e$ | 18.5 in.$^2$ |
| Total weight flow rate, $\dot{m}g$ | 7.8 lb/sec |
| Oxidizer flow rate, $\dot{m}_0 g$ | 5.72 lb/sec |
| Fuel flow rate, $\dot{m}_f g$ | 2.08 lb/sec |
| Combustion volume, $V_c$ | 272 in.$^3$ |
| Chamber area ratio, $f_c/f_t$ | 5.85 |
| Chamber diameter, $d_c$ | 5.25 in. |
| Chamber length, $l_c$ | 11 in. |

The design information concerned with the injector and the coolant ducts for this motor are discussed in SEC. **30**.

### 30. Propellant Injection

The purpose of the "injector" is to introduce the propellant into the combustion chamber in such a manner that mixing or atomization takes place quickly and uniformly with minimum reduction in fuel pressure. For a monopropellant, the injector orifices are designed to produce a fine spray, facilitating atomization and evaporation. For a bipropellant, the orifices are designed to produce impinging high-velocity streams, facilitating mixing.

If the pressure drop through the injection orifices is too low, irregular combustion and even acoustic oscillations may result. On the other hand, a very high pressure drop is undesirable because it requires heavy pressurizing tanks. A compromise which is satisfactory for acid–aniline is to have the dynamic head $q[=\frac{1}{2}\rho v^2]$ of the impinging streams such that $q \geqslant 8640$ lb/ft$^2$, where $\rho$ is liquid density and $v$ is velocity. In the case of liquid in turbulent flow through the short tubular orifices of a typical injector, the pressure drop $\Delta p$ is related to the dynamic head $q$ by the equation

$$\Delta p = Kq, \tag{83}$$

where $K$ is a dimensionless orifice coefficient; its value is between 1.2 and 2.0, depending upon the shape of the orifices and degree of turbulence. The rate of heat transfer to the combustion

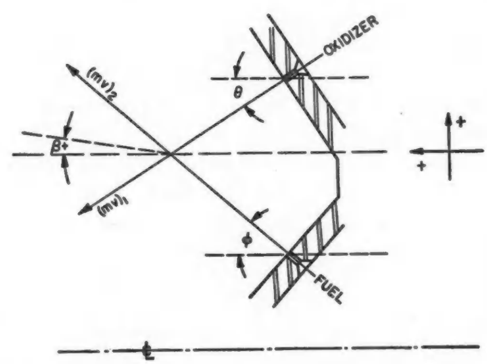



FIG. 25. Diagram of the resultant momentum of the stream produced by two impinging injector streams.

chamber walls is very sensitive to the geometric configuration of the entering propellant streams. Since regenerative cooling of rocket motors is a difficult problem, the orientation of the jet must be very carefully controlled. Variations of a few degrees in the orientation of these jets may cause local variations in heat transfer of 50 to 100 percent, and may be responsible for the failure by melting of the chamber wall, since no great excess of cooling is possible.

*The impinging-jet injector.*—Figure 24 shows the construction of a typical acid–aniline injector with impinging-jet mixing. This is one of the simplest of many possible arrangements of orifices. The orifices are made removable, so that their hydraulic contour and orifice coefficients may be carefully controlled. Minimum heat transfer to the chamber is secured with this injector when the direction of the resultant momentum of the streams after impingement is almost parallel to the axis of the cylindrical combustion chamber. A typical orientation of the streams is shown in Fig. 25. The orientation of the resultant streams may also indirectly affect the performance—that is, the characteristic velocity, $c^*$—of the motor by changing the entrance temperature of the fuel coolant. Figure 26 shows the appearance of water flowing through a small impinging-jet injector. The data for the injector of Fig. 24, used in the 1500-lb thrust motor described in SEC. **29** are as follows: number of pairs of jets, 8; fuel orifice diameter, 0.096 in.; oxidizer orifice diameter, 0.1285 in.; nominal $\Delta p$ for fuel jets, 67 lb/in.$^2$; nominal $\Delta p$ for oxidizer jets, 100 lb/in.$^2$; angle of resultant momentum $\beta$, $+5°$ (outward). In this case the jets are designed for unequal pressure drops because the hydraulic circuit of the fuel line included a pressure drop in the motor coolant duct.

*Other types of injector.*—An interesting bipropellant injector in which the fluids impinge in an annulus rather than at discrete points is shown in Fig. 27. This injector produces somewhat better mixing and combustion efficiency than the impinging-jet type, by allowing a converging cone of fluid to intersect a diverging cone. It seems also to produce a higher rate of heat transfer to the motor walls.

Injectors whose primary function is atomization rather than mixing are built quite differently, as may be seen in Fig. 28, which shows an injector used with the monopropellant nitromethane in a 200-lb thrust motor. Here the liquid is whirled in a spray head, so that upon emergence it is broken into fine droplets by centrifugal action. The rate of heat transfer from this type of injector is considerably less sensitive to orientation than that from the bipropellant types.

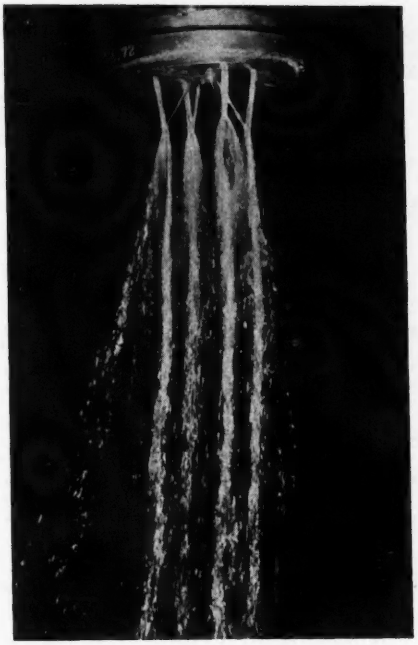


FIG. 26. Water test of the orifice alignment of a 200-lb thrust injector.

The art of injector design is as yet largely empirical, since little is known about the internal processes in a liquid rocket motor. Direct measurements of velocity, temperature, density and chemical composition are very difficult because of the high temperatures of the reactants.

### 31. Initiation of Combustion

Starting liquid rocket motors presents a number of serious problems not encountered in conventional power plants. A warm-up period at rated thrust is not feasible since the high rate of propellant consumption makes each unused second of operation extremely costly in loss of impulse-weight ratio. Moreover, combustion must begin promptly (within a few tenths of a

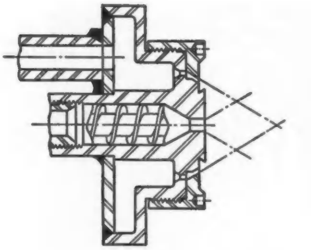


FIG. 27. Section of a bipropellant injector producing two intersecting conical sheets of nitric acid and aniline.

second), else the delayed ignition of a chamber filled with accumulated propellants may cause a destructive pressure surge, or "hard start."

With propellants that ignite spontaneously upon contact, such as acid and aniline, safe ignition is secured by limiting the initial flow to a fraction of the normal full flow until combustion is well established. These particular components react first at their liquid interfaces, releasing heat at a rate proportional to the surface area or degree of mixing. This heat is dissipated at a rate that depends upon the state of agitation of the liquids and the shape and temperature of the chamber into which they are injected. If the heat-flow balance is unfavorable, the ignition may delay until excess propellant has accumulated, resulting in a "blow." It is essential to have the hydraulic system so adjusted that both propellant components arrive *simultaneously*, and that the mixture ratio during the initial transient flow not deviate too far from the stoichiometric value. An initial flow of $\frac{1}{10}$ to $\frac{1}{6}$ normal will usually establish in 2 or 3 sec sufficient combustion pressure that the transition to full flow may be effected in 1 or 2 sec more. One successful starting technic has been to insert rupture disks in the hydraulic lines which burst when the propellant feed pressure has reached a specified small fraction of its final value. The delay in building up to full pressure then provides the necessary partial flow.

*Initiation of combustion with catalysts.*—Hydrogen peroxide may be promptly decomposed if there is injected with it about 3 percent by weight of saturated calcium permanganate catalyst solution. Sometimes the permanganate is directed against a target to assist in atomizing it. If a fuel is being combined with peroxide, the catalyst flow may be stopped within a second or two and the reaction will continue. This has proved to be a very reliable method of ignition. A typical catalyst-peroxide-fuel injector is shown in Fig. 29. Peroxide decomposition may also be initiated by passing it through a bed of catalyst stones.

*Initiation of combustion with sparks and flames.*—Nonspontaneous propellants, such as nitromethane and the various liquid-oxygen combinations, must be ignited in the vapor phase by a spark or flame. Consequently the ignition device must be placed where it will not be flooded by the initial flow of propellant; for example, note the position of the spark plug in Fig. 28. Spark plugs are convenient for small motors because repeated starts may be made, although they have the disadvantage that the electrodes burn off in a relatively short time. For large motors such as the V-2, a pyrotechnic device in the motor interior is used to secure greater heat release, as otherwise there is risk of quenching the ignitor.

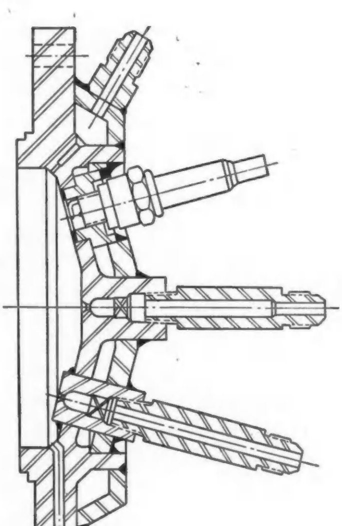


FIG. 28. Section of a monopropellant injector designed to produce a finely atomized spray by centrifugal action. Note spark plug for ignition.

Nitromethane has the unfortunate property that a pressure surge, or "hard start," resulting from delayed ignition (or thermal decomposition in coolant ducts) may initiate a detonation wave in the nitromethane which can be propagated along the fuel supply line and even into the supply tank. For this reason the spark-plug circuits used with nitromethane are equipped with cut-offs that terminate the current after a fixed maximum safe delay interval measured from the instant when fluid begins to enter the chamber. It is interesting to note that nitromethane is difficult to ignite unless gaseous oxygen is present, although oxygen may be discontinued as soon as the combustion begins. Detonation traps have been developed that stop the progress of the wave beyond a certain point in the supply line by shattering the line and dispersing its contents prior to the arrival of the wave.

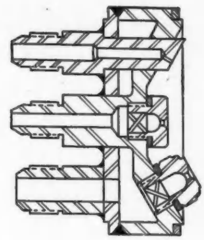


FIG. 29. Section of a bipropellant injector (peroxide-nitromethane) in which ignition is accomplished by a preliminary reaction between a permanganate catalyst and hydrogen peroxide.

#### 32. Heat Transfer in Rocket Motors

(*a*) *Magnitudes of thermal quantities.*—A rocket motor operates under more severe conditions of high temperature and of continuous rate of heat release than any other heat engine. For these reasons problems of heat transfer are among the most acute and important in rocket-motor design. Motors may be broadly classified into two categories, depending on whether the heat transferred from the high-velocity hot gases is absorbed by the motor materials (uncooled type) or whether all or part of the propellant is used to absorb the heat (cooled type). The latter type can be further subdivided into the kind in which the coolant liquid absorbs heat as it circulates in ducts around the motor and is then injected into the combustion chamber (regeneratively cooled type), and the kind in which a part of the coolant liquid is injected directly into the motor in such a way as to provide a coolant film on the inner wall surface (film-cooled type).

It is not at once evident whether the flowing propellant can absorb the heat transmitted through the walls and still remain in the liquid phase. Measurements of the total heat transmitted show that it is often possible to absorb this heat without boiling the coolant. A marked scale effect exists for cooling. The rate of coolant flow increases linearly with motor thrust, while the motor area to be cooled increases less rapidly. This means that in certain critical cases a given propellant can be used to cool a large motor but not a small one.

A description of the numerical magnitudes of the thermal quantities involved in typical regeneratively cooled motors may be of value in providing orientation. The heat of combustion of the red fuming nitric acid (RFNA)–aniline propellant, for example, is approximately 1800 Btu/lb. At sea level, with $p_c = 300$ lb/in.$^2$, less than 50 percent of this heat is converted into kinetic energy of the jet, and almost the entire remaining portion remains in the jet in the form of thermal energy. In the motor, between 2 and 3 percent of the heat due to combustion passes through the chamber and nozzle walls into the coolant, which returns it again to the combustion chamber when regenerative cooling is utilized. This heat transmission is brought about by a density of heat flow $q$ of about 1.0 Btu/in.$^2$ sec in the motor chamber and 2.5 to 3.0 Btu/in.$^2$ sec in the nozzle throat, which is the position of maximum heat transmission. The hottest industrial furnaces have a rate of heat transfer only about one tenth of that encountered in the rocket motor. In Fig. 30 is shown the distribution of heat flow density $q$ along the axis of a 200-lb thrust motor made of an aluminum chamber and a copper nozzle block.

The heat transmitted to the motor can be reduced by a factor of two to four by inserting refractory liners. However, refractory materials so far developed have a limited lifetime in rocket motors.

For the RFNA–aniline propellant the temperature of the chamber gases $T_G$ is about 4500°F,

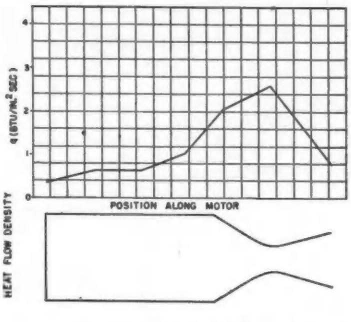



FIG. 30. Plot of heat transfer density $q$ as a function of position along the rocket motor axis.

falling to 3000° in the nozzle throat and to 2000°, or less, at the exit section of the exhaust nozzle. Most available metals melt well below these temperatures. The fact that uncooled motors of even short duration can be built is caused by the existence of a large temperature drop across the boundary layer between the main gas stream and the motor walls. The equilibrium temperature $T_{WG}$ of the inner wall surface (next to the hot gases) in regeneratively cooled motors ranges from 800° to 1400°F for steel alloys. This temperature is determined largely by the wall thickness (0.1 to 0.2 in. for steel alloys) and the thermal conductivity of the material. The temperature of the wall next to the coolant liquid $T_{WL}$ is usually limited to a value below the boiling point of the liquid, which, for example, is 650°F for aniline at 500 lb/in.². Another steep temperature drop occurs across the boundary layer between the main body of the coolant liquid and the cooled side of the motor wall. This "film drop" amounts to about 200°F. A typical value of the mean liquid temperature $T_L$ after absorbing the transmitted heat is 300°F.

The relationship of these various temperatures is illustrated in Fig. 31, and the magnitudes of the various thermal quantities described are listed in Table VI.

The temperature of the motor wall for a given temperature of the chamber gases is controlled principally by the temperature of the coolant next to the wall. For this reason it is desirable to have the "film-drop" $T_{WL}-T_L$ as low as possible. The thermal conductance $h$ of this film, which is defined as the heat transfer density per unit temperature difference across the film, increases with the velocity of the coolant, and the film-drop may vary from 30 or 40 to several hundred Fahrenheit degrees. Consequently it is desirable to have the coolant velocity as high as possible, consistent with an acceptable pressure drop across the coolant duct system. In present designs the coolant velocities lie between 15 and 50 ft/sec. The velocity is usually increased in the vicinity of the nozzle throat. An effort is made to limit the total pressure drop through the cooling ducts to about 50 lb/in.².

Since only 2 to 3 percent of the heat liberated in the chamber is transmitted through the walls, and since even this small percentage is absorbed by the coolant with but a scant margin of safety, any factor that affects the heat transfer is of great importance in the design of a motor. As has been discussed in SEC. **30**, the heat transfer is very sensitive to small variations in the flow pattern of the propellant inside the chamber. It has been found experimentally that slight deviations in dimensions during fabrication of the injector are capable of affecting both the performance parameter $c^*$ and the density of heat flow $q$ by factors of 20 to 100 percent.

A promising means for controlling the heat flow to the wall of a rocket motor, first used by the Germans in the V-2, is the technic of introducing small quantities of a liquid at many points distributed uniformly over the interior surface. The liquid so introduced is spread over the wall in a thin film and eventually evaporated.

TABLE VI. Typical regeneratively cooled rocket-motor thermal quantities. RFNA-aniline propellant, aniline coolant.

| | |
|---|---|
| Propellant heat of combustion | 1800 Btu/lb |
| Fraction appearing as jet kinetic energy | 40 percent |
| Fraction appearing as jet enthalpy | 60 percent |
| Fraction transmitted through walls | 3 percent |
| Chamber density of heat transfer, $q_c$ | 1.0 Btu/in.² sec |
| Nozzle density of heat transfer, $q_n$ | 2.5–3.0 Btu/in.² sec |
| Temperature of gases in chamber, $T_G$ | 4500°F |
| Temperature of gases in nozzle throat | 3000°F |
| Temperature of gases in nozzle exit | 2000°F |
| Temperature of motor wall on gas side, $T_{WG}$ | 1000°F |
| Temperature of motor wall on coolant side, $T_{WL}$ | 500°F |
| Coolant film drop, $T_{WL}-T_L$ | 200°F |
| Mean coolant temperature, $T_L$ | 300°F |
| Coolant boiling temperature at 500 lb/in.² | 650°F |

The essential advantage of this method, termed "film cooling," is that the screening film of coolant fluid is permitted to vaporize, thus increasing its heat-absorbing capacity many fold over that of a system in which the fluid must remain in the liquid phase. It has the further advantage that the heat does not have to be conducted *through* the motor walls, resulting in a great saving in strength and weight. In an ideal film-cooled motor the motor walls would require no external cooling and would never reach a temperature higher than the boiling point of the coolant.

A logical extension of the film-cooling process is to increase the number of cooling orifices indefinitely; in other words, use porous walls. The coolant fluid then oozes in uniformly at all points of the surface. Such a technic is sometimes referred to as "sweat cooling."

The coolant film may be produced either by part or all of the propellant or by some additional liquid which may be inert (for example, water) or may contribute energy to the combustion process. A notably successful example of the use of film cooling is the German V-2 rocket motor, in which about 3 percent of the total mass flow is diverted to the alcohol cooling film. About half of this takes part in combustion and hence is not wasted.

(*b*) *Mechanism of heat transfer.*—The hot gases transmit heat to the exposed motor walls principally by convection and radiation. Direct conduction may be neglected. Theoretical calculations[20] of the heat transfer indicate that up to 30 percent of the heat absorbed by the chamber walls is caused by radiation. In the nozzle, where temperatures and dimensions are smaller, radiation is not important. This situation may alter if it becomes possible to use significantly higher temperatures in the combustion chamber.

The important quantity of heat transmitted by convection to the chamber wall is difficult to calculate accurately. The density of convective heat flux $q_c$ through the hot gas boundary layer, or film, is proportional to the temperature difference across the film; that is,

$$q_c = h_G \Delta T, \tag{84}$$

where $h_G$ is called the *gas-film coefficient*, or *thermal conductance*, and may, for example, be expressed in terms of 1 Btu/in.² sec °F as the unit; $\Delta T$ is $T_c - T_{WG}$, where $T_c$ is the chamber-gas temperature and $T_{WG}$ the inner-wall temperature. If it is assumed that the same conditions of convective heat transfer exist in the chamber as in a long straight pipe—a very doubtful assumption—then one may use an equation due to von Kámán[21] based on the analogy between fluid friction and heat transfer, to calculate $h_G$. The equation is

$$h_G = q_c/\Delta T = C_H c_p \rho_c v_c, \tag{85}$$

where $c_p$ is the specific heat at constant pressure; $\rho_c$, the mass density of combustion products; $v_c$, the velocity of chamber gases; $C_H$, a dimensionless heat-transfer coefficient. Use of the equation of continuity reduces Eq. (85) to the more practical form

$$h_G = C_H c_p (\dot{m}/f_c), \tag{86}$$

which indicates the manner in which heat transfer depends upon mass-flow rate and chamber cross-sectional area $f_c$. The coefficient $C_H$ is a function of flow conditions, Reynolds number and surface roughness. It varies greatly from one section of the chamber to the next. A rough calculation[21] gives a typical value for $C_H$ of 0.0022, a figure that can be in error by as much as 100 percent.

The heat flux density is determined primarily by the conductance $h_G$, since its value is much less than those of the other conductances in the heat flow path, in a manner analogous to the way in which the current in a series electric cricuit is determined by a predominately large resistor. A typical value of $h_G$ is 0.0003 Btu/in.² sec °F.

In practical calculations of rocket motor-wall temperatures, it is necessary to have prior empirical knowledge of $q$, the heat flux density. If this number is known, one can proceed stepwise to calculate all the critical temperatures of Table VI in the following sequence.

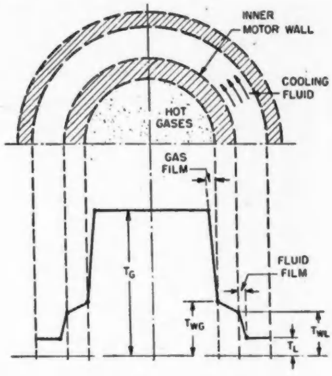



FIG. 31. Plot of temperature levels across a section of rocket motor chamber. Note the steep gradients in the gas and liquid "films" next to the solid wall.

[20] These calculations involve a number of doubtful assumptions about the emissivity of the hot gases and their velocity distribution in the combustion chamber.

[21] Th. von Kármán, "The analogy between fluid friction and heat transfer," *Trans. Am. Soc. Mech. Engrs.* **61**, 705–710 (1939).

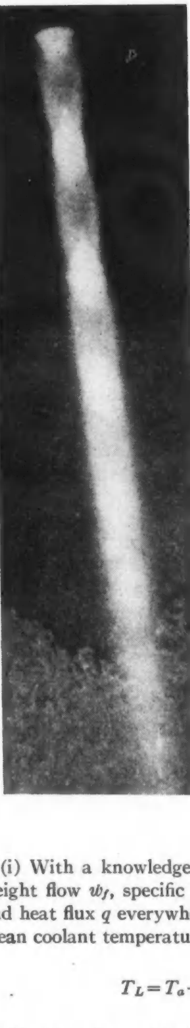


FIG. 32. The exhaust jet of a 200-lb thrust rocket motor, showing the luminosity striations due to oblique shock waves.

(i) With a knowledge of the chamber area $A$, coolant weight flow $w_f$, specific heat $c_p$, ambient temperature $T_a$ and heat flux $q$ everywhere over the surface, calculate the mean coolant temperature $T_L$ from the equation

$$T_L = T_a + (1/w_f c_p)\int_0^A q dA. \tag{87}$$

(ii) At various critical sections of the motor, such as the nozzle throat and the region of maximum $T_L$, assume a coolant velocity and calculate the *liquid-film conductance* $h_L$, which then permits calculating the temperature $T_{WL}$ of the chamber wall-liquid interface from the equation

$$T_{WL} = T_L + q/h_L. \tag{88}$$

If this temperature is too far above the local boiling point of the liquid at the estimated local pressure, it may be necessary to choose new values of velocity, area or weight flow rate. A typical value of $h_L$ is 0.005 Btu/in.$^2$ sec °F.

The calculation of $h_L$, the liquid-film thermal conductance, is made with the aid of a semi-empirical formula obtained by methods of dimensional analysis. A typical formula of this kind is the following:[22]

$$(h_L D/k) = 0.023(vD\rho/\mu)^{0.8}(\mu c_p/k)^{0.4}, \tag{89}$$

where $h_L$ is the film conductance for a circular tube; $D$, the diameter of the tube; $k$, the thermal conductivity of the liquid; $\rho$, the mass density of the liquid; $v$, the velocity of the liquid; $\mu$, the viscosity of the liquid; $c_p$, the specific heat of the liquid. The terms in parenthesis are all dimensionless quantities. From left to right, they are known as Nusselt, Reynolds and Prandtl numbers, respectively. Certain quantities in this equation are not accurately known for rocket-propellant liquids; for example, thermal conductivity and viscosity. A correction to $h_L$ must be made for the effect of bending the duct into circular arcs of rather small radius. This correction is not accurately known, but may be as much as 25 to 50 percent.

(iii) Estimate the necessary thickness $d$ of chamber or nozzle wall to withstand operating stresses at estimated operating temperatures; then, knowing the thermal conductivity $K$ of the wall material, calculate the temperature $T_{WG}$ of the gas-wall interface by using the customary equation for heat flow through a slab. The nozzle throat will normally be a critical region. The equation is

$$T_{WG} = T_{WL} + qd/K. \tag{90}$$

If $T_{WG}$ is such that the material is too weak for the stresses involved, then a new choice of thickness, material or value of $T_{WL}$ must be made and the calculations repeated.

(*c*) *The coolant pressure drop.*—Since feed pressure influences the critical parameter of over-all weight, it is necessary to keep the pressure drop through the coolant ducts as low as possible. If the coolant velocities chosen in the preceding subsection result in too high a pressure, new velocities must be chosen and the calculations repeated. Care must be exercised in choosing the dimensions of the coolant ducts and in their subsequent fabrication, since at constant weight flow rate the pressure drop through a circular coolant tube, for example, varies inversely as the fifth power of the tube diameter!

Bending the cooling duct into a helix of moderate radius affects the flow in such a manner as to increase the pressure drop by as much as 30 percent. Using the best available hydraulic information, the pressure drop in the 1500-lb thrust motor described in SEC. **29** was computed to be 40 lb/in.$^2$, a figure that was later verified experimentally.

### 33. Characteristics of Rocket Jets

(*a*) *Shock waves.*—A striking feature of the jet issuing from a rocket nozzle is the presence of

[22] W. H. McAdams, *Heat transmission*, (McGraw- Hill, 1942), p. 168.

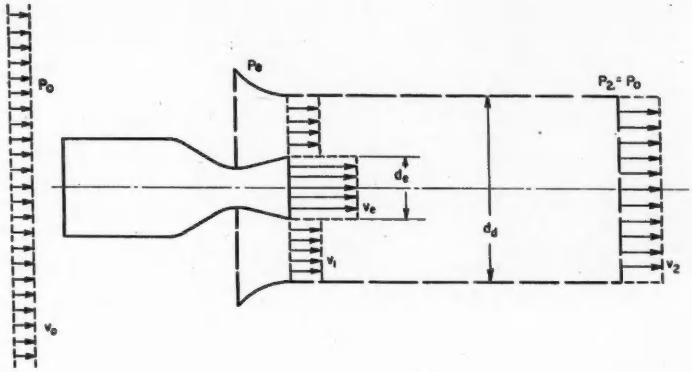



FIG. 33. An augmentor tube for enhancing jet thrust. This device increases the momentum change that can be produced by the kinetic energy of the primary jet.

clearly defined oblique shock waves[23] numbering up to six or more, such as can be seen in Fig. 32. These consist of regions in which sudden, almost discontinuous, changes in pressure, density, velocity and entropy occur in the flowing gas. They remain fixed relative to the nozzle, and have no effect upon thrust unless they occur *inside* the nozzle, which will be the case only for a markedly over-expanded nozzle. It has been shown by Prandtl[24] that the spacing $d$ (in.) of these luminous shock waves is related to the thrust $F$ (lb) by the simple equation

$$d=(F/3)^{1/2}. \tag{91}$$

The *over-all* length $l$ (ft) of a typical rocket flame may be roughly estimated from the equation

$$l=(F/10)^{1/2}. \tag{92}$$

In the case of an acid–aniline motor the after-burning or luminous flame may be virtually eliminated by the addition of 6 percent of $KNO_3$ to the acid.

(*b*) *Augmentation.*—When a typical rocket jet is permitted to escape at atmospheric pressure, kinetic energy is wasted which might be more efficiently used. If this jet be permitted to entrain surrounding air, part of the energy can be used to impart a directed velocity to additional mass and thus supply increased momentum or thrust. A tube, surrounding the rocket, in which this mixing may occur is called an "augmentor" (Fig. 33).

Analysis indicates that up to 35 percent thrust augmentation is possible with this type of tube if the system is stationary. Augmentation decreases rapidly if the system is in motion, falling to half the stationary value when the system is moving, relative to the atmosphere, with 5

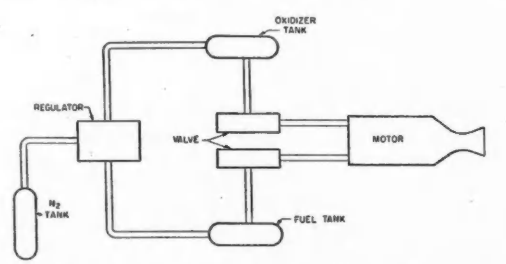



FIG. 34. Typical hydraulic circuit of a liquid rocket system using compressed gas for propellant pressurization.

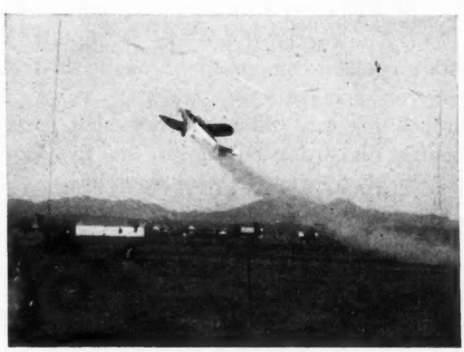


FIG. 35. Light airplane taking off with the assistance of small solid propellant rockets. Note large angle of climb.

[23] See reference 10, Vol. III, pp. 213–222; also A. Stodola, *Steam and gas turbines* (McGraw-Hill, 1927) vol. I, pp. 83–94; vol. II, pp. 1006–1016.

[24] L. Prandtl, "Stationary waves in a gas stream," *Physikal. Zeit.* 5, 599 (1904).

FIG. 36. Typical liquid-propellant assisted take-off rocket power plant of approximately 1000 lb thrust. It was designed for permanent mounting in the engine nacelle of an aircraft, and was built by the Aerojet Engineering Corporation.

percent of the rocket exhaust velocity. In practice the necessary cylinders would be rather bulky to incorporate in an operating model.

## Liquid-Propellant Rocket Systems and Their Applications

### 34. Basic Components

A complete, self-contained, liquid-propellant rocket system must include not only the motor but flow-control devices, a propellant supply in suitable containers, and a means of pressurizing the propellant. Figure 34 shows a typical gas-pressurized system. A strong tank holds nitrogen at a pressure of 2000 lb/in.². During operation, this supply pressure falls to about 600 lb/in.². The nitrogen flows through a regulator which reduces the pressure to a constant value of 500 lb/in.². The regulated pressure is applied over both propellant components through a hydraulically or pneumatically operated valve and through one-way check valves which prevent possible disastrous mixing of the two liquid components by way of the nitrogen line.

The pressurized liquids are released to the rocket motor by a hydraulically operated valve (or in some cases by rupture diaphragms) placed as close to the motor as possible, thus facilitating the simultaneous arrival of the two fluids in the chamber. In a field-service type of system, the propellant valve is arranged to open simultaneously with the nitrogen valve, which then permits flow of liquid during a gradual increase of nitrogen pressure in the ullage above the liquids and consequent smooth build-up of combustion pressure in the rocket motor. It is interesting to note in this connection that the monopropellant nitromethane must be kept from flowing too rapidly into the motor injector manifold, or adiabatic heating of the air-vapor mixture compressed in the manifold may cause an explosion.

The fabrication of the pressurized propellant and gas vessels demands the best technic possible, since they must have minimum weight and be able to withstand high stresses with only a small factor of safety.

### 35. Feed-Pressure Technic

For a short-duration system, the use of a tank of pressurizing gas at ambient temperatures and at pressures between 2000 and 3000 lb/in.² is quite common. If it be assumed that no heat is transferred to the gas as it expands from an initial pressure $p_0$ to a final pressure $p_f$ in a supply tank of volume $V_0$, and that the gas passes through a regulator and fills propellant tanks of volume $V_p$ at a regulated pressure $p_r$, then analysis shows that these quantities are related by the equation

$$V_0/V_p=\gamma p_r/(p_0-p_f), \tag{93}$$

where $\gamma$ is the ratio $c_p/c_v$ of specific heats of the pressurizing gas. In actual practice enough heat is absorbed during the operating period (about 1 min) that $\gamma$ must be replaced by a smaller "effective" value $\gamma'$, which is empirically determined. Thus for nitrogen $\gamma$ is 1.40, and $\gamma'$ may be as low as 1.25.

A very significant saving in weight may be achieved if the pressurizing gases can be generated as they are used from a chemical reaction. This results from the fact that the container for the chemicals will be very small and need withstand only feed pressure rather than many multiples of feed pressure. Furthermore, since the gases are usually hot when generated, they may be used at an absolute temperature several times the normal ambient temperature of bottled gases, thus reducing the density and mass of gas required by the same factor. Gases may be generated conveniently from rocket propellants themselves, either liquid of solid. A considerable body of technic remains to be developed, however, before this valuable weight-saving method comes into practical use.

As the duration of operation of a rocket increases, the tank weight necessary to contain bottled pressurizing gas also increases. Above some critical duration of about 1 min (which is less the larger the rocket thrust), pressurization can be accomplished with a smaller weight penalty by using turbine-driven centrifugal pumps than by using stored gases.

It is not possible here to go into the design details of these pumping plants. The gas turbine can be conveniently driven at about 10,000 rev/min by using 2 or 3 percent of the rocket propellant in a special decomposition chamber as a gas generant. The detailed technical problems of developing pumps to handle fluids such as nitric acid and liquid oxygen are numerous. Small pumps have been driven by turbines that project directly into the rocket exhaust and are rotated windmill fashion. It should be pointed out here that whereas chemical generation of pressurizing gases reduces the weight of the gas container it does not reduce the weight of the propellant container. With a pump feed system the propellant container becomes a light-weight low-pressure vessel.

### 36. Assisted Take-Off and Superperformance of Airplanes

The greatest demands upon an airplane power plant are made during take-off, at which time it must accelerate a maximum load to flying speed in a limited time and simultaneously overcome air or water drag. Extra thrust from a rocket motor is very valuable during this period.

The following examples show typical improvements in performance when the thrust of an aircraft is increased 30 to 50 percent by means of rockets.

(i) The take-off distance with normal load may be decreased to about two-thirds normal value. This is of value on carriers and small airfields.

(ii) The load may be increased 20 percent with normal take-off distance. This is of value on long flights where extra fuel is needed, or under emergency overload conditions.

(iii) The take-off distance at an altitude of 10,000 ft is reduced from nearly twice the sea-level distance without rockets to approximately the sea-level value. This is of value on high-altitude airfields.

(iv) The rate of climb at any altitude may be almost doubled. This is of value as a safety factor immediately after take-off, or in escaping pursuit, overtaking an enemy and avoiding antiaircraft fire.

(v) The speed of level flight may be increased about 25 percent. This is useful for the reasons mentioned in (iv).

Figure 35 shows a small airplane taking off with 150 lb of jet assistance. The increase in angle of climb over the usual angle is quite obvious. Another application of JATO (jet-assisted take-off) is in the take-off and landing of "supersonic" ultra high-speed airplanes, which may have such small wing surfaces as to be unable to take off or land without rocket assistance and braking action. The rockets used may be solid (Fig. 16) or liquid (Fig. 36), depending upon their duration, and may be droppable or permanently installed in the airplane. Both types

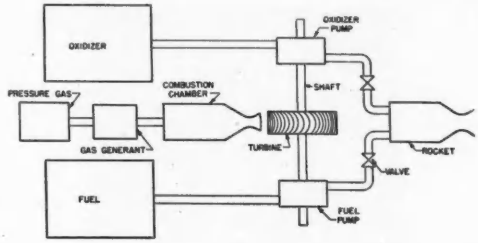



FIG. 37. Diagram of a rocket system—the "turborocket"—in which the liquid propellants are pressurized by a turbine driven pump.

have been developed commercially. A typical weight break-down of a liquid-rocket JATO system designed to give 1300 lb thrust for 60 sec is as follows: propellants, 400 lb; pressurizing gas, 25 lb; tanks, 120 lb; plumbing and structure, 75 lb; rocket motor, 50 lb; total, 670 lb.

### 37. Propulsion of Aircraft

The rocket is well suited to the propulsion of aircraft at very high altitudes and at speeds above that of sound, in the region where propellers are not effective. The duration of operation is long enough that only pump-fed liquid rocket systems are feasible. This duration may range typically from 10 min to 1 hr.

Rocket propulsion will be especially useful for the "flying laboratory" type of airplane designed

TABLE VII. Typical V-2 data.

| | |
|---|---|
| Maximum range | 200 mi (approx.) |
| Maximum altitude (vertical trajectory) | 100 mi (approx.) |
| Maximum velocity, at end of burning | 5000 ft/sec |
| Length over-all | 46 ft |
| Body diameter | 5.5 ft |
| Total weight, including propellant | 27,300 lb |
| Empty weight, excluding warhead | 5100 lb |
| Warhead weight | 1700 lb |
| Propellants: | |
| Fuel | 75% ethyl alcohol, 25% water |
| Oxidizer | Liquid oxygen |
| Mixture ratio, oxygen to fuel | 1.25 |
| Rocket motor: | |
| Nominal full thrust (sea level) | 55,000 lb |
| Duration of thrust | 80 sec |
| Throat diameter | 16 in. |
| Exit diameter | 29 in. |
| Combustion pressure | 225 lb/in.$^2$ |
| Thrust coefficient | 1.30 |
| Specific impulse (sea level) | 202 sec |

FIG. 38. A German V-2 rocket being serviced at White Sands Proving Ground, New Mexico, preparatory to a night firing.

to study the aerodynamics of transonic, supersonic and even hypersonic flight. Rockets known as "rocket airfoil testers" (RAFT) have already been sent aloft with small airfoils attached to the nose, supersonic aerodynamic data being sent back from them by radio to ground recorders.

One pump-fed system, designated the "turborocket," is shown in Fig. 37. In it the liquid propellant is pressurized by means of pumps driven by a gas turbine. The gas turbine may be operated by the products of combustion of the same propellants as are used in the rocket motor, or, as in the German V-2 rocket, by decomposing a separate fluid such as hydrogen peroxide. If rocket propellants are used, it is sometimes necessary to add a diluent to reduce the gas temperature resulting from the reaction to a value that will not damage the turbine blades.

A turborocket system was used by the Germans in their rocket fighter Me-163B. This power plant provided 3300 lb thrust during the climbing period of 3 min and by means of a throttling system provided 300 to 400 lb thrust during cruising. It achieved top speeds approaching sonic velocity and was limited to a total flying time of 10 to 20 min. At take-off, approximately 50 percent of the weight of the plane was propellant, the combination used being hydrogen peroxide and "*C*-Stoff."

Another spectacular German rocket aircraft, known as the "Natter," was launched *vertically* from a pole, climbing to 40,000 ft in approximately 1 min. It then released a simultaneous barrage of two dozen or so rockets at an approaching bomber formation. In the next few moments the nose of the aircraft was mechanically detached by the pilot and discarded, and a web-type parachute was released from the tail. The resulting deceleration effected separation of the pilot from the cockpit, and he completed the landing *via* his own parachute, alone and unaccompanied by any of the original expendable aircraft. It would be a blasé pilot indeed who would find such an assignment dull!

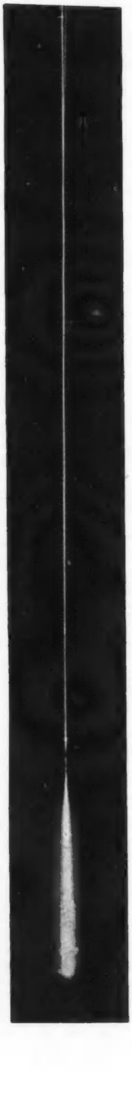


FIG. 39. Night firing of liquid-propellant sounding rocket with solid-propellant booster, showing brief dark interval between termination of booster and initiation of missile combustion.

TABLE VIII. The WAC-Corporal sounding rocket.

| | |
|---|---|
| Maximum vertical altitude | 43.5 mi |
| Over-all length | 16 ft |
| Body diameter | 1.0 ft |
| Total weight, including propellant | 691 lb |
| Empty weight, less payload | 272 lb |
| Payload | 25 lb |
| Propellant: | |
| Fuel | Aniline with 20% furfural |
| Oxidizer | Red fuming nitric acid |
| Thrust (sea level) | 1500 lb |

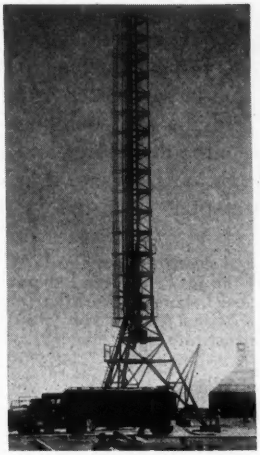


FIG. 41. Launching tower at White Sands Proving Ground from which the WAC Corporal sounding rocket was launched. View of missile-proof observing room in background. Rockets have descended vertically from 40 mi to land close to this structure.

### 38. Missiles and Sounding Rockets

The Germans developed a great variety of liquid rocket missiles, ranging from an anti-aircraft projectile 4 in. in diameter and 60 in. long, called "Taifun," to the massive V-2, weighing nearly 14 tons (Fig. 38). It is not feasible to attempt to describe them all here. The famous—or infamous—V-2 will suffice to illustrate the nature of these missiles. It is propelled by liquid oxygen and ethanol-water, and is stabilized during the early stages of its flight by graphite vanes which project into the exhaust flame. The rocket motor, which is made of ordinary mild steel about $\frac{1}{4}$ in. thick, has several rings of small holes through which a small fraction of the fuel component of the propellant is admitted to form a protective cooling film over the inner wall of the combustion chamber. The structure of the V-2 is shown in Fig. 21, and some quantitative data about it are tabulated in Table VII.

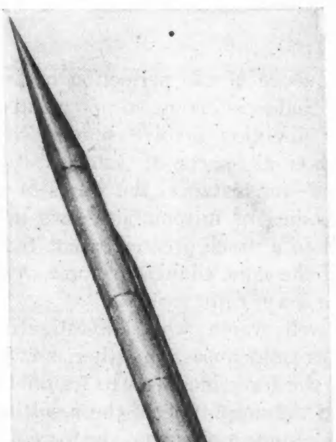


FIG. 40. The WAC Corporal sounding rocket, which reached 43 mi altitude.

The first American rocket to achieve altitudes comparable to the V-2 was developed in our laboratory and is named the WAC Corporal. This rocket is relatively small, carrying a payload of 25 lb, and was designed primarily for sounding measurements in the upper atmosphere. It carried no jet vanes for stabilization, but depends upon the aerodynamic stability of its fins to keep it on a vertical course. It is launched from a 100-ft vertical tower with a solid-propellant (ballistite) booster rocket originally developed for the Navy under the name of "Tiny Tim." This rocket was modified to deliver a 50,000-lb thrust for 0.5 sec. Such initial launching is necessary in order that the WAC Corporal

achieve an aerodynamically stable velocity before leaving the launcher. Figure 39, a night picture taken of the launching process, shows cessation of the large booster flame just as the missile flame begins and subsequent separation of the flight paths of booster and missile. The structure of the WAC is shown in Fig. 40, and certain information[25] is listed in Table VIII. Figure 41 is a view of a WAC launching tower.

The authors gratefully acknowledge their indebtedness for much of the information in these two papers to numerous members of the staff of the Jet Propulsion Laboratory.[26]

***(The third, and final, paper in this series will deal with two as yet undeveloped fields—escape from the earth and the application of nuclear energy to rocket propulsion.)***

[25] For a general description of the WAC Corporal and the activities of the Jet Propulsion Laboratory of Guggenheim Aeronautical Laboratory, see the July 1946 issue of *Engineering and Science* (published by the Alumni Association, California Institute of Technology).

[26] A large part of the information here presented has been abstracted from *Jet propulsion*, a book prepared by this staff for the Army Air Forces; this volume is in process of being rewritten in an unclassified version.

# On the Significance of Sudden Variations of Force

ERNA M. J. HERREY
*Queens College, Flushing, New York*

ELEMENTARY physics teaches that a force acting upon a spring produces a displacement that is proportional to the force. However, this is true only if the force is applied very slowly. Otherwise, vibrations in the natural frequency of the spring are superimposed on the static displacement. The amplitude of these vibrations increases with the suddenness with which the force is applied so that the maximum dynamic displacement may exceed by far the static displacement due to the same force.

The same phenomenon is known in electricity where it is put to use in the square-wave testing of communication circuits when the response to the sudden application of an emf reveals important characteristics of the circuit. It is also known in acoustics, where it is called "shock excitation." This is, for instance, the cause of the report when a shell leaves the barrel of a rifle or a gun, or when a stopper is removed from a champagne bottle, whereby the inside pressure is suddenly reduced.

In the field of mechanical vibrations this effect has lately assumed considerable significance in consequence of the perfection of high-speed vehicles.[1] Sudden starting, stopping and changing speed or direction involve sudden changes of force. Thus all parts of vehicles capable of vibrations—for instance, the wings of airplanes or the bodies of automobiles—are apt to be displaced to a much greater extent than would be true if the same changes of force were taking place at a lower rate.

It is well worth while investigating what constitutes suddenness, or rather, how the time in which the force increases to its final value is related to the amplitude of the resulting vibrations. We assume at first that the force $f$ increases to its final value $F$ according to an exponential law,

$$f=F(1-e^{-t/\alpha}). \qquad (1)$$

The constant $\alpha$ indicates how fast the main increase of force takes place. For instance, $f$ reaches 90 percent of its final value in the time

[1] Compare P. Le Corbeiller, "A classical experiment illustrating the notion of 'jerk,'" *Am. J. Physics* **13**, 156 (1945).

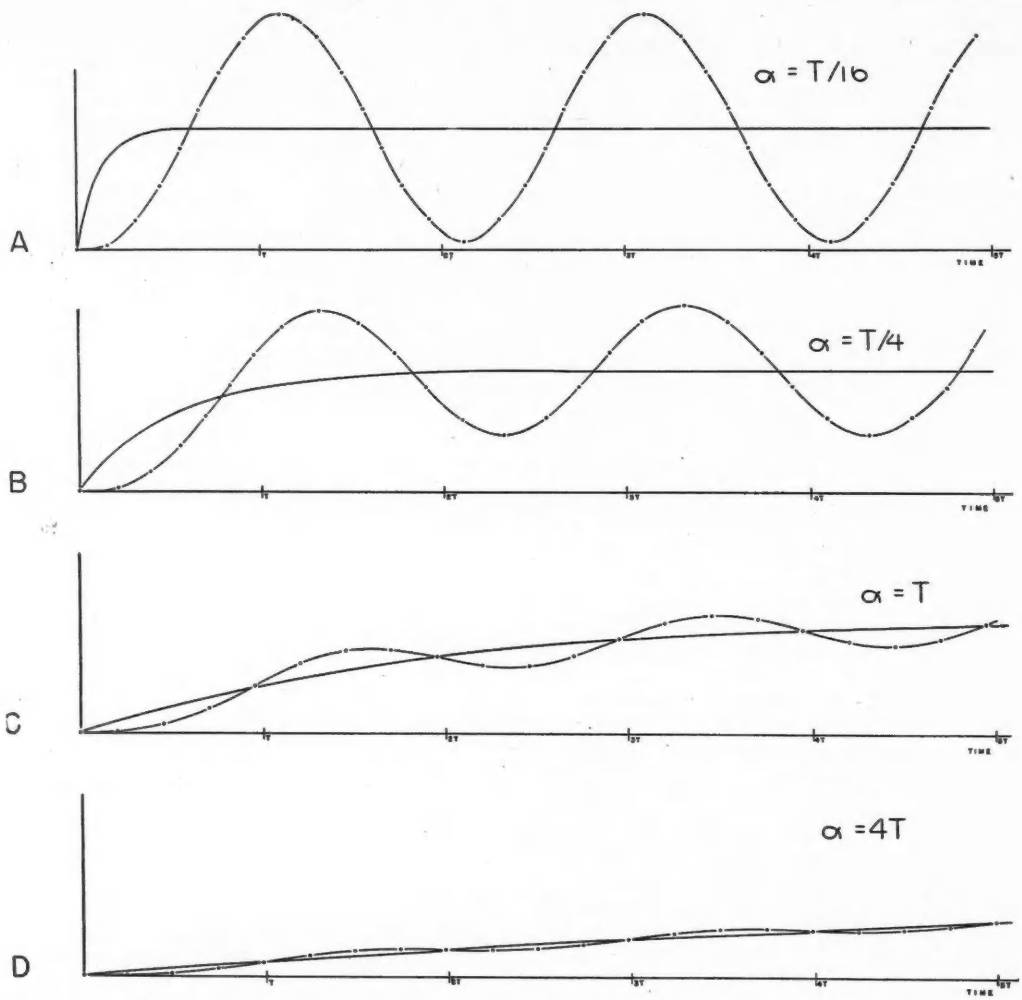



FIG. 1. Displacement *vs.* time. *Dashed curves*, motion produced by force $F(1-e^{-t/\alpha})$; *solid curves*, static displacements.

$2.3\alpha$. If damping is neglected, the differential equation for the resulting dynamic displacement $x$ is

$$m(d^2x/dt^2)+sx=F(1-e^{-t/\alpha}), \quad (2)$$

where $m$ is the mass and $s$ the stiffness of the system. With the initial conditions $x=0$ and $dx/dt=0$ for $t=0$, and with the period of free vibrations $T$ substituted for $2\pi(m/s)^{\frac{1}{2}}$, the solution is

$$x=(F/s)(1-\cos^2\epsilon\, e^{-t/\alpha}) -(F/s)\sin\epsilon\,\sin(2\pi t/T+\epsilon), \quad (3)$$

$$\tan\epsilon=T/2\pi\alpha. \quad (3')$$

A graphical representation of this solution for various values of the rate of growth $\alpha$ is given in Fig. 1, dashed curves. The first term on the right-hand side of Eq. (3) follows the same course as the static displacement $x_0=f/s$ shown by the solid line; it increases with time and is somewhat larger than $x_0$ for small values of $t$, but it gradually approaches $x_0$ and its maximum value $(x_0)_{max}=F/s$ for large $t$. The second term is a sinoidal function in the period of free vibration $T$

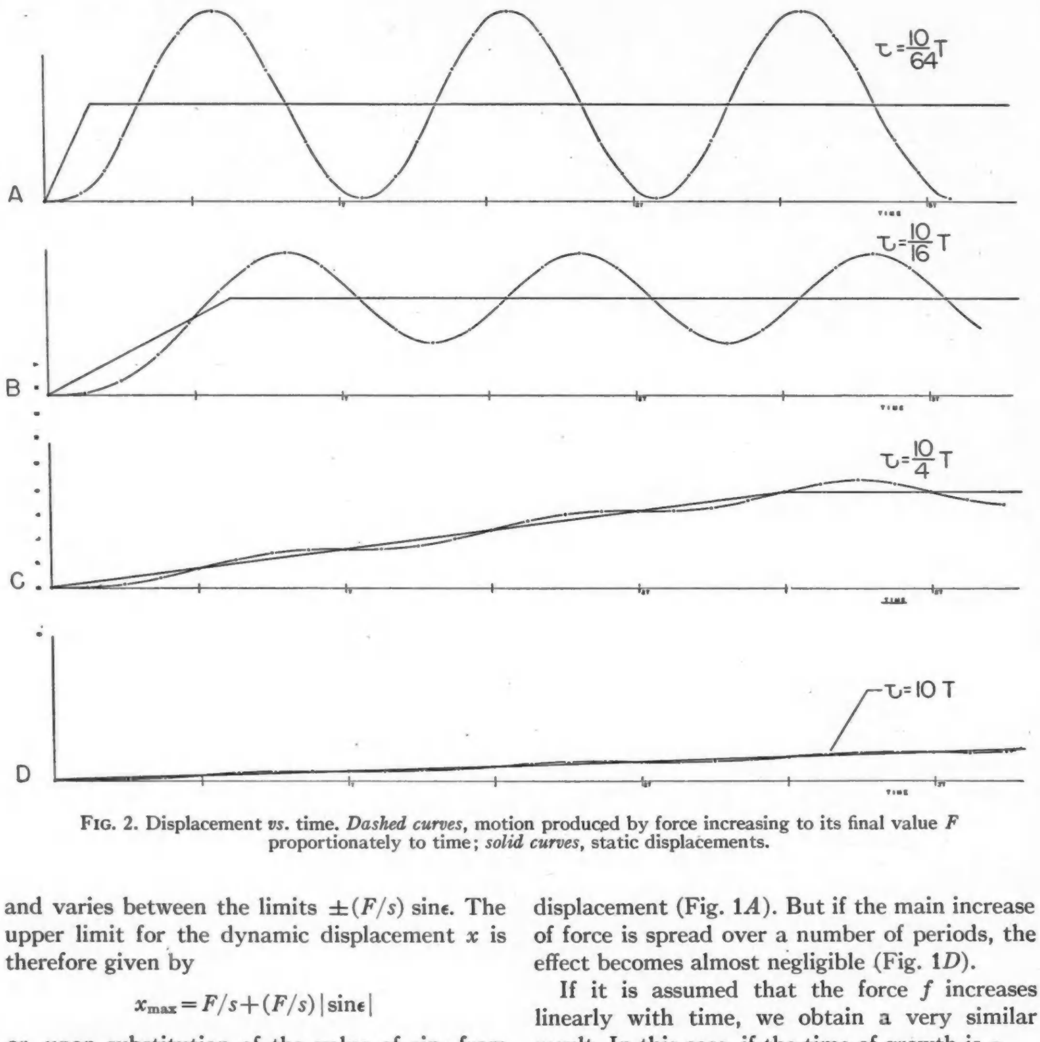



FIG. 2. Displacement *vs.* time. *Dashed curves*, motion produced by force increasing to its final value $F$ proportionately to time; *solid curves*, static displacements.

and varies between the limits $\pm(F/s)\sin\epsilon$. The upper limit for the dynamic displacement $x$ is therefore given by

$$x_{\max}=F/s+(F/s)\,|\sin\epsilon|$$

or, upon substitution of the value of $\sin\epsilon$ from Eq. (3′), and of $(x_0)_{\max}$ for $F/s$,

$$x_{\max}=(x_0)_{\max}[1+(1+(2\pi\alpha/T)^2)^{-\frac{1}{2}}]. \quad (4)$$

This shows that the maximum dynamic displacement exceeds the maximum static displacement by a factor depending on $\alpha/T$ only, that is, depending only on the relation of the growth of force to the period of free vibrations. If $\alpha$ is very small compared with $T$, or if the main increase of force takes place during a fraction of a period, the maximum dynamic displacement is almost twice as great as the maximum static displacement (Fig. 1*A*). But if the main increase of force is spread over a number of periods, the effect becomes almost negligible (Fig. 1*D*).

If it is assumed that the force $f$ increases linearly with time, we obtain a very similar result. In this case, if the time of growth is $\tau$,

$$f=(F/\tau)t, \quad \text{for } 0\leqq t\leqq\tau, \quad (5a)$$

$$f=F, \quad \text{for } t\geqq\tau. \quad (5b)$$

The differential equation of motion becomes

$$m(d^2x/dt^2)+sx=(F/\tau)t, \quad \text{for } 0\leqq t\leqq\tau, \quad (6a)$$

$$m(d^2x/dt^2)+sx=F, \quad \text{for } t\geqq\tau. \quad (6b)$$

The initial conditions for Eq. (6a) are $x=0$, $dx/dt=0$ for $t=0$; and for Eq. (6b) they are that, for $t=\tau$, $x$ and $dx/dt$ must be equal to the solution of Eq. (6a). It follows that

$$x=(F/s)(t/\tau)-(F/s)(T/2\pi\tau)\times\sin(2\pi t/T),\quad \text{for } 0\leqq t\leqq\tau, \tag{7a}$$

$$x=(F/s)-(F/s)\left[\frac{\sin(\pi\tau/T)}{\pi\tau/T}\right]\times\cos(2\pi(t-\tfrac{1}{2}\tau)/T),\quad \text{for } t\geqq\tau. \tag{7b}$$

The first term in each solution represents the static displacement $f/s$, while the second term represents a vibration in the free period $T$ and with an amplitude $(F/s)(T/2\pi\tau)$ for $0\leqq t\leqq\tau$ and $(F/s)\sin(\pi\tau/T)/\pi\tau/T$ for $t\geqq\tau$. These amplitudes are also the maximum differences between dynamic displacement $x$ and static displacement $x_0$. A graphical representation of the motion for various values of the time of growth $\tau$, is shown in Fig. 2, dashed curves. Since solution (7a) for $t\leqq\tau$ is a function that never decreases (its first derivative is never negative), and since for $t=\tau$ the solutions (7a) and (7b) are equal, the maxima of solution (7b) represent the maximum dynamic displacements at any time. Substituting $(x_0)_{\max}$ for $F/s$, we get

$$x_{\max}=(x_0)_{\max}\left(1+\left|\frac{\sin(\pi\tau/T)}{\pi\tau/T}\right|\right). \tag{8}$$

As in the previous case, the relation of the time of growth $\tau$ to the free period $T$ is the determining factor for the magnitude of the dynamic displacement. Again, if $\tau$ is very small compared with $T$ (Fig. 2*A*) the maximum dynamic displacement approaches twice the value of the maximum static displacement. If $\tau$ is very large compared with $T$ (Fig. 2*D*), dynamic and static displacements differ very little from each other. For intermediate values of $\tau$, however, the maximum dynamic displacement also depends on the phase of the vibration at the instant when the final value $F$ of the force is reached. If the phase at which this occurs happens to be zero or if $\tau$ is a whole multiple of $T$, the sine term vanishes. This means that for $t\geqq\tau$ the vibratory motion ceases and dynamic and static displacement are equal. But if $\tau$ is an odd multiple of $\frac{1}{2}T$, as in Fig. 2*C*, the amplitude of the vibratory motion is relatively great.

One could of course investigate this phenomenon by substituting other laws of increase of force besides the linear and the exponential ones treated here. However, for the extreme cases of very sudden and of very slow increase of force—sudden and slow relative to the period of free vibration $T$—one would always get the same results: $x_{\max}=2(x_0)_{\max}$, and $x_{\max}=(x_0)_{\max}$, respectively. In Fig. 3 the excess of maximum dynamic over maximum static displacement in percentage of the maximum static displacement is plotted as a function of the time of growth $\tau$ for the two laws of increase of force considered here. Since, for an exponential increase of force the time of growth is theoretically infinitely great, the time in which 90 percent of the final force is reached has been used instead ($\alpha=\tau/2.3$). It is seen that—apart from the rather accidental effect of the phase of the vibration at the moment when the force becomes constant, which leads to periodic changes in the lower curve—the two curves are almost identical.

These effects of sudden changes of acceleration are familiar to passengers in high-speed vehicles. Of course some protection is provided through the action of shock-absorbing springs. But shocks not completely absorbed elsewhere create discomfort that is related to the phenomenon just discussed: perception of motion takes place in the inner ear where, in the semicircular canals, a liquid is displaced in proportion to the acceleration. Abrupt changes in acceleration should bring about displacements up to twice the static values, and thus the sensation of an acceleration in excess of its actual value may be created. The consequence is the familiar feeling of nausea due to a peculiar reflex connection between the nerves originating in the semicircular canals and those supplying the gastro-intestinal system. This would explain why particular discomfort is experienced in elevators when they start and stop, and in swings. In either case rates of change of acceleration are high.

It is interesting to note in this connection that some animals, in reacting to stimulation of the nerves ending in the semicircular canals, exhibit certain reflex movements such as spreading of the toes, getting ready to jump and other muscular reactions. Most of these reflexes are observed at the beginning and again at the end of the motion, that is, when the changes of acceleration are highest. All reflexes disappear after extirpation of the labyrinth containing the semicircular canals.[3]

Although these physiological effects are readily observed, the effects upon vehicles in which we travel are of course of still greater consequence. This is because provisions must be made in the

[3] *Handbuch der Hals-Nasen-Ohrenheilkunde* (Springer, 1926), vol. 6, p. 1037.

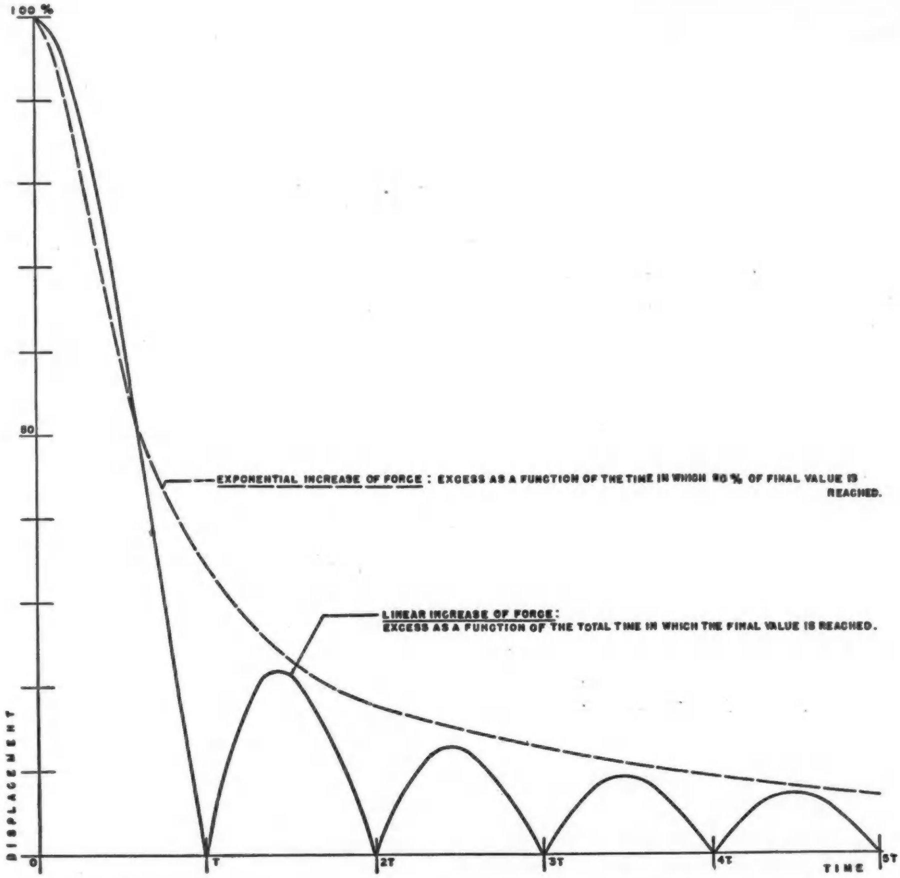



FIG. 3. Excess of maximum dynamic over maximum static displacement in percentage of static displacement as a function of the time of growth.

structure of vehicles and, where surface vehicles are concerned, in the design of tracks or roads not only for our comfort but for our safety. Obviously the points where sudden changes of force are inevitable, as in the transition from a straight section of highway to a curve, require particular attention. Let us consider here how this problem in highway design is met by inserting transition spirals so as to prevent too sudden an increase of centripetal force. The task is mainly to determine the necessary transition length.

On unbanked curves the centripetal force on the chassis is provided solely by side friction between tires and road surface. Since the maximum useful coefficient of side friction is about[4] 0.3, the maximum possible centripetal acceleration at any speed is about 10 ft/sec$^2$. Since the body of the automobile is not rigidly connected to the chassis but linked to it by springs, it tends to stay in a straight path even if side friction forces the chassis into the curve. Consequently the body is deflected sidewise relative to the chassis. This lateral angular deflection, or "body roll," must increase to such a value $\Theta$ that the restoring forces set up in the springs

[4] Moyer, "Skidding characteristics of automobile tires on road surfaces and their relation to highway safety," Bull. 120, Iowa Eng. Exp. Sta. (Iowa State College, 1934); Dietz and Harling, "Die Fahrlage des Kraftwagens in der Kurve," *Deut. Kraftfahrforsch.* **44**, 1 (1940).

become large enough to provide the necessary centripetal force on the body.[5] If road curvature is introduced too suddenly the body roll may become excessive, up to twice its static value, leading to the danger of overturning.

The necessary minimum time of growth of centripetal force $\tau$ can be ascertained from Fig. 3 and the length $L$ of the transition spiral determined from $L=v\tau$, $v$ being the maximum expected speed on that highway. Of course it is impossible to adjust the length of the transition spiral so that the vibration of every car on that road section is at zero phase just when it arrives at the circular section of the curve. Speeds and periods of the various cars on the road being different, it is necessary to use the maxima of the lower curve; that is, $\sin(\pi\tau/T)$ in Eq. (8) must be replaced by its maximum value, 1. Substituting the roll angle $\theta$ for the linear displacement $x$, we get from Eq. (8),

$$\theta_{\max}-(\theta_0)_{\max}=(\theta_0)_{\max}\frac{\sin(\pi\tau/T)}{\pi\tau/T}\leqq(\theta_0)_{\max}(T/\pi\tau).$$

The static roll angle $\theta_0$ is a linear function of the centripetal acceleration $a$ and varies between $0.1a$ and $0.6a$ degrees for different cars.[6]

The period of free lateral vibration $T$ varies between 0.6 and 1 sec.[7] Therefore, even in the most unfavorable cases,

$$\theta_{\max}-(\theta_0)_{\max}=0.6A/\pi\tau<0.2A/\tau,$$

where $A$ is the final centripetal acceleration on the circular section, and $A/\tau$ is the average rate of change of acceleration, for which the symbol $C$ is used in highway engineering literature. Thus,

$$\theta_{\max}-(\theta_0)_{\max}<0.2C. \tag{9}$$

It follows that in order to keep the excess body roll within safe limits, $C$ should not be greater than about 5 times the maximum allowable excess body roll. The corresponding length $L$ of the transition spiral can be found from the relation $C=A/\tau=v^3/RL$ (since $A=v^2/R$ and $\tau=v/L$). So far no attempt has been made to calculate $C$ from the properties of the automobile as has been tried here. Experimental determinations of the appropriate length of transition spirals are also scarce and contradictory.[8] In present-day highway design practice calculations of transition curves are based on a rather arbitrary empirical value of[9] $C=2$ or 3 ft./sec.[3]

This analysis may help to clarify the physical significance of the rate of change of acceleration $C$ and to determine its safe maximum value from the mechanical constants of the vehicle. From Eq. (9) it seems quite possible that a higher value of $C$ could safely be adopted. If the design of transition spirals were based, for instance, on $C=5$ ft/sec$^3$, the excess body roll due to transition from tangent into circular highway sections, and vice versa, would still in no case exceed 1° and usually would be much less. Of course, there would be considerable simplification and saving in time and cost if the design of transition spirals could be based on a higher value for $C$, leading to lower values for the transition length $L$. The safe maximum value for $C$ is also of some consequence for the determination of distance and space required for overtaking and furthermore for the process of "weaving" from one lane into another. In both cases the minimum distance depends on the maximum safe curvature of the vehicle's path as well as on the necessary transition length between straight and curved path.

However, for the actual design of transition spirals it is necessary to know not only whether a given length of transition spiral satisfies conditions of safety as regards body roll and danger of overturning, but also whether it is mechanically possible for the car to turn in fast enough so as to stay in its proper lane on the road without skidding. Especially on sharp curves, where speeds have to be low, the time needed to develop the so-called slip-angle—the angle between the direction of travel and the plane of the wheels necessary to provide the centripetal force of side

[5] On a banked curve all or part of the centripetal force on the chassis as well as upon the body is provided by gravity. The body roll $\theta$ is then proportional to $a-eg$, where $e$ is the banking angle in radians. Thus the considerations for unbanked curves apply also to banked curves if $a$ is replaced by $a-eg$.

[6] Stonex and Noble, "Curve design and tests on the Pennsylvania Turnpike," *Proc. Highway Research Board* **20**, 449 (1940).

[7] Schilling and Fuchs, "Modern passenger car ride characteristics," *J. Applied Mechanics* **8**, A59 (1941).

[8] Warren and Hazeldine, "Experimental transition curves," *J. Inst. Municipal and County Eng.* **65**, 1021 (1939); Orchard, Correspondence to "A survey of the present position in road transition curve theory," *J. Inst. Civil Eng.* **12**, 464 (1939); "Clover-leaf loops," *J. Inst. Civil Eng.* **23**, 210 (1944/45).

[9] Moyer, reference 4, esp. p. 113; Barnett, "Transition curves for highways," U. S. Dept. Agric., Bur. Public Roads, Washington, 1938; Agg, *Construction of roads and pavements* (1940).

friction—may exceed the time $\tau$ necessary to avoid excessive body roll. These circumstances and the maximum rate at which the steering wheel can be turned may possibly have a bearing on the required lengths of transition spirals where narrow curves and low speeds are concerned. A study of these questions is in progress. Yet it seems from these considerations that transition spirals can be considerably shorter than is commonly assumed at the present time.

The effects incident to change of force have been discussed here for cases that happen to be of some practical importance. However, it seems important to draw attention to some general conclusions that apply to any vibratory system, whether mechanical, acoustic or electric, and that may be summarized as follows:

(1) If a force is applied very suddenly, its effect may be as much as twice that produced by the same force when it is applied slowly.

(2) The magnitude of the effect does not depend on the absolute value of the time of growth or decrease but rather on the ratio of this time to the period of free vibration of the system.

(3) If the system is almost completely rigid ($s\rightarrow\infty$), the period of free vibration is almost zero, and any time of application of force may be considered as large. In this case the rate of change of force has no appreciable influence on the behavior of the system.

# A Rigorous *WLT* System of Dynamics Based on the Mach Law of Inertia and the Law of Gravitation

AUSTIN J. O'LEARY
*The City College, New York 10, New York*

THERE are two distinct systems of dynamics in use throughout the world: a *mass-length-time*, or *MLT*, system, and a *weight-length-time*, or *WLT*, system. The latter, used mainly by engineers, is generally regarded as a purely practical system and, to my knowledge, has not previously been set up on a rigorous basis. It is my purpose to restate the basic laws of dynamics and to develop a *WLT* system that is an exact equivalent of the *MLT* system. It will be seen how the two systems can be built in the same fashion, step by step, upon a common foundation.

## Two Basic Laws

*The Mach law of inertia.*[1,2]—The Mach law is here stated for an isolated system of any number of particles in a way that gives a new meaning to the relations between two types of acceleration and two corresponding types of force. The actual time-rate of change of velocity of a particle will be called the *effective* acceleration of the particle, in distinction from what I shall call its *individual* accelerations in its interactions with other particles. Measurement indicates that the effective acceleration of each particle is equal to the vector sum of all its individual accelerations, where, in each interaction, the two particles of an interacting pair are taken to have individual, oppositely directed, central accelerations in a given ratio to each other, independently of their interactions with other particles. Let $\mathbf{a}_x$ denote the effective acceleration of any particle $x$ in an isolated system of $n$ particles; $\mathbf{a}_{x-y}$, the individual acceleration of particle $x$ in its interaction with any other particle $y$; and $\mathbf{a}_{y-x}$, the corresponding individual acceleration of particle $y$ at the same instant; then we have

$$\mathbf{a}_x=\sum_{y=1}^{n}\mathbf{a}_{x-y}, \quad (y\neq x) \tag{1a}$$

$$\mathbf{a}_{y-x}/\mathbf{a}_{x-y}=-k_{x-y}=-1/k_{y-x}, \tag{2a}$$

where the reciprocals $k_{x-y}$ and $k_{y-x}$ are positive numerics; or, for any arbitrary interval of time,

$$\Delta\mathbf{v}_x=\sum_{y=1}^{n}\Delta\mathbf{v}_{x-y}, \quad (y\neq x) \tag{1b}$$

$$\Delta\mathbf{v}_{y-x}/\Delta\mathbf{v}_{x-y}=-k_{x-y}=-1/k_{y-x}, \tag{2b}$$

[1] E. Mach, *Science of mechanics* (Open Court, 1942), esp. pp. 264–306.
[2] R. B. Lindsay and H. Margenau, *Foundations of physics* (Wiley, 1936), esp. pp. 85–97.

where $\Delta\mathbf{v}$ with one subscript denotes effective change in velocity and with two subscripts denotes individual change in velocity. Since the individual accelerations and individual changes in velocity derive their physical significance from Eq. (1) along with Eq. (2), these two sets of equations are complementary; in general, neither one has physical meaning apart from the other. Letting $s$ denote any third particle of an isolated system, measurement reveals further that

$$k_{x-y}=k_{x-s}/k_{y-s}, \tag{3}$$

a relation for which there is no *a priori* expectation from Eqs. (1) and (2).

Equations (1)–(3) may be closely related to experience in a projectile experiment (Fig. 1)[3] in which pairs of telescoping tubes are successively projected apart in a horizontal direction by a spring as they fall through the air. With the earth taken as the frame of reference, the vector sum of the individual accelerations of each tube in its interactions with the particles of the earth is a specific known quantity $\mathbf{g}$. Thus, for the motion of two tubes $x$ and $y$ as they fly apart, Eq. (1a) takes the form $\mathbf{a}_x=\mathbf{a}_{x-y}+\mathbf{g}$, $\mathbf{a}_y=\mathbf{a}_{y-x}+\mathbf{g}$, whence the individual acceleration $\mathbf{a}_{x-y}$ is readily discernible as the horizontal component of $\mathbf{a}_x$, and $\mathbf{a}_{y-x}$ as the horizontal component of $\mathbf{a}_y$; and since the tubes start from rest, $k_{x-y}$ is equal to $d_y/d_x$, the ratio of the horizontal distances traveled by the tubes in falling through equal vertical heights. In effect, we attain isolation of the two interacting tubes without having resorted to the unsatisfactory concept of isolated pairs of particles in stating the Mach law. The constancy of $\mathbf{g}$, which makes possible the isolation, will be discussed later.

[3] A. J. O'Leary, *Am. J. Physics* **14**, 120 (1946). In the original arrangement of the projectile experiment, three tubes $A$, $B$ and $C$ were so constructed that $A$ and $C$ could each be compared with $B$, and $B$ plus $C$ with $A$.Tubes have since been constructed to demonstrate Eqs. (1)–(3) more directly. Three tubes $x$, $y$, and $s$ for the purpose are shown in Fig. 1. Tube $s$ fits snugly inside tube $x$, tube $x$ inside tube $y$, and tube $s$ is fitted inside tube $y$ by moving adapting rings $R$ to appropriate positions along the length of tube $s$. Two pins $P$ and $P'$ fitting lengthwise slots in tubes $y$ and $s$ prevent rotation of one tube inside the other in each of the three combinations. A sliding collar $I$ on each tube serves as a position indicator (see Fig. 2 of the earlier paper). For convenience in detecting the existence of Eq. (3) and checking the inertial measurements quickly on a beam balance in classroom demonstrations, the masses of tubes $s$, $x$ and $y$ were made equal to 50, 70 and 100 gm, respectively, so that $k_{x-s}=1.40$, $k_{y-s}=2.00$ and $k_{x-y}=0.70$.

An obvious simplification follows from Eq. (3). We select a particular particle $s$ as a standard with respect to which $k_{x-s}$ may conceivably be measured for each particle $x$. Substituting $k_{x-s}/k_{y-s}$ for $k_{x-y}$ in Eq. (2) and comparing the different expressions for any particle $x$ in its interactions with various particles $y$ ($y=1, 2, 3, \cdots$), we see that the numeric $k_{x-s}$ is characteristic of particle $x$ relative to particle $s$. We therefore replace the numerics $k$ by a primary physical quantity, assigning 1 unit to particle $s$ and $k_{x-s}$ unit to particle $x$. An original investigator would naturally follow a definite pattern in setting up Eqs. (2) and (3). If he happened to start with the ratios $\mathbf{a}_{x-s}/\mathbf{a}_{s-x}$ (here designated by the symbol $k_{s-x}$) and $\mathbf{a}_{y-s}/\mathbf{a}_{s-y}$ (here designated by the symbol $k_{s-y}$) along with the ratio $\mathbf{a}_{y-x}/\mathbf{a}_{x-y}$, getting $k_{x-y}=k_{s-y}/k_{s-x}$ in place of Eq. (3), the numeric characteristic of any particle $x$ relative to the standard particle $s$ would be $k_{s-x}$ in place of $k_{x-s}$; and following the usual procedure, the investigator would end up with "inverse mass" as the primary quantity. However, before reporting his findings, he would probably change to a primary quantity in direct proportion to volume and weight of particular materials, rather than inverse proportion. We shall therefore proceed from Eq. (3) as written in its present form.

The primary quantity is called *mass* in the *MLT* system. By definition, the mass of the International Kilogram is 1 *kilogram mass*, or 1 kgm [≡1000 gm≡$10^6$ mgm, etc.], and the mass of particle $x$ is $k_{x-s}$ kgm. Letting $m$ denote mass, we have $m_x=k_{x-s}m_s$. Secondary quantities in the *MLT* system are defined in terms of length, time and mass, each unit of mass constituting a third physical dimension in the units of secondary quantities. A like procedure is followed in the *WLT* system presented here; to conform with engineering practice as closely as possible, the primary quantity will be called *standard weight* in this system. By definition, the standard weight of the International Kilogram is 1 *kilogram weight*, or 1 kgwt, [≡1000 gwt≡$10^6$ mgwt, etc.], and the standard weight of particle $x$ is $k_{x-s}$ kgwt. Like mass, standard weight is a scalar quantity;[4]

[4] Since *local weight* is a vector quantity, there may be some question as to the fitness of the name "standard weight" for a scalar quantity. On the other hand, there is a precedent for such usage in the traditional use of the terms

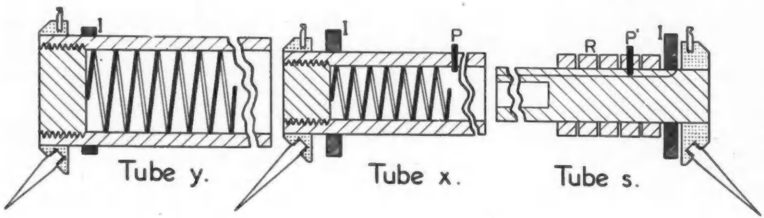



FIG. 1. Cross section of three telescoping tubes used in a projectile experiment to demonstrate Eqs. (1)–(3).

direction is not involved in any way in its definition. Letting $w$ denote standard weight, we have $w_x = k_{x-s} w_s$. Secondary quantities in the *WLT* system are defined in terms of length, time and standard weight, each unit of standard weight constituting a third physical dimension in the units of secondary quantities. However, there is a difference in the two procedures. The definitions of secondary quantities in the two systems differ in such a way that the kilogram mass and the kilogram weight have different physical dimensions. This academic distinction between mass and standard weight, and between the kilogram mass and kilogram weight, will be considered later. Meanwhile, the symbol $m'$ will be used to denote either mass or standard weight throughout the remainder of this section so as to make the discussion applicable to both systems of dynamics.[5] Equations (2) and (3) may now be combined, yielding

$$m'_x \mathbf{a}_{x-y} = -m'_y \mathbf{a}_{y-x}, \tag{4a}$$

or,

$$m'_x \Delta \mathbf{v}_{x-y} = -m' \Delta \mathbf{v}_{y-x}. \tag{4b}$$

A specific assumption is contained in the Mach law, namely that the mutual interactions in a group of particles are all independent of one another. Aside from other limitations, the whole structure of dynamics depends upon the applicability of this assumption. We happen to know from quantum mechanics that closely packed particles of atomic size do not behave as individual particles interacting independently with one another. So the Mach law does not apply to every known case even when expressed in relativistic form. Nonetheless, Eqs. (1) and (4) are confirmed in a wide range of applications and may rightly be said to express a law.[6]

*Frames of reference.*—In his writings and lectures, Mach strongly disavowed the Newtonian concept of space.[7] He interpreted motion to mean

*atomic weight* and *molecular weight*. Direction is not associated with either of these quantities; for example, the atomic weight of oxygen is 16 grams, not 16 grams directed downward. The names atomic weight and molecular weight are out of place in the *MLT* system, but they fit nicely into the scheme of the *WLT* system. The atomic weight of an element is the standard weight of $N_0$ atoms of the element, where $N_0$, the Avogadro number, is the number of atoms contained in 16 grams of oxygen. All things considered, the name standard weight would seem to be the most appropriate one for the third primary quantity in the *WLT* system.

[5] It might be well to revive Newton's original term *quantity of matter* to mean either mass or standard weight (as here defined, with no implication concerning the numbers of elementary particles constituting a body) in instances where one does not wish to connote a particular dimension. Some such added name for the primary quantity would be useful in many statements meant to apply to both systems of dynamics. For units having no dimensional connotation, we might use the names *kilogram* and *pound* without the qualifying term *mass* or *weight*; for example, a 5-kilogram body would mean one for which $m = 5$ kgm and $w = 5$ kgwt.

It is more appropriate to call the International Kilogram a standard particle or standard quantity of matter than to call it a standard of mass. It serves as a standard for the physical quantity *mass* neither more nor less than it does for the physical quantity *standard weight.*

[6] A physical law does not mean a relation determined for all time. Laws are made provisional by the very nature of measurement. An "absolute proof" of any law is meaningless because of limitations in the precision of measurements, the change in conditions often produced by the very fact that a measurement is made, the impossibility of performing observations under all conceivable circumstances, and the impossibility of making experiments conform to the idealized conditions under which laws are assumed to obtain. Measurement cannot prove laws but can disprove them, or perhaps better, can reveal exceptions to the accepted rule. Definitions and laws are developed in the course of time through successive stages of approximation. We are here concerned with one stage in the development.

[7] Some indication of the Mach view may be gained from the following quotation (reference 1, p. 281): "When we say that a body $K$ alters its direction and velocity solely through the influence of another body $K'$, we have asserted a conception that it is impossible to come at unless other bodies $A$, $B$, $C$, $\cdots$ are present with reference to which the motion of the body $K$ has been estimated. In reality, therefore, we are simply cognizant of a relation of the body $K$ to $A$, $B$, $C$, $\cdots$. If now we suddenly neglect $A$, $B$, $C$, $\cdots$ and attempt to speak of the deportment of the body $K$ in absolute space, we implicate ourselves in a twofold error. In the first place, we cannot know how $K$ would act in the absence of $A$, $B$, $C$, $\cdots$; and in the second place, every means would be wanting of forming a judgment of the behavior of $K$ and of putting to the test what we had predicated—which latter therefore would be bereft of all scientific significance." See also Sections 7 and 8, pp. 286–289, and elsewhere in reference 1.

that celestial bodies move, not relatively to space, but only relatively to all the other bodies in a finite universe. More specifically, he affirmed that an effectively isolated particle has zero acceleration with respect to the center of mass of all the other particles in the universe; similarly for the center of mass of an effectively isolated system of particles. In this interpretation of motion, inertia, as well as gravitation, depends upon a kind of mutual action between bodies. Many subtle implications[8] are contained in this view that are beyond the scope of this paper. However, independently of the abstruse problem of inertial frames of reference, the Mach formulation is no less valid than any other approach leading to the same equations;[9] it is only our interpretation of the equations of mechanics that depends on the way in which inertial frames are specified. The Mach method is a vast improvement over that of Newton in economy of thought, and it frees mechanics from the intuition and anthropomorphism inherent in Newton's laws and definitions.

*Law of gravitation.*—It is found that the individual acceleration of any particle $x$ in a gravitational interaction with any other particle $y$ is given by the equation

$$\mathbf{a}_{x-y} = G m_y' \mathbf{r}_{y-x} / r_{y-x}^3, \tag{5}$$

where $G$ is a proportionality factor, and $\mathbf{r}_{y-x}$ is the position vector of particle $y$ relative to particle $x$. The local gravity acceleration $\mathbf{g}$ of a particle with respect to the earth is the resultant of a large number of individual accelerations; for a particle $x$ at a given point, we have

$$\mathbf{g} = G \sum (m_y' \mathbf{r}_{y-x} / r_{y-x}^3) - \mathbf{a}_c - \mathbf{a}_C,$$

where the summation extends over all particles $y$ composing the earth, $\mathbf{a}_c$ is the centripetal acceleration and $\mathbf{a}_C$, the Coriolis acceleration at the given point. This expression for $\mathbf{g}$ does not involve $m_x'$. Thus, it follows from Eqs. (1) and (5) that $\mathbf{g}$ at any given point should be the same for every particle independently of its physical state, chemical constitution, or other variable condition. The well-known experiments of Galileo, Newton, Bessel, Southerns, and Eötvös have demonstrated that such is the case within small limits of error for all the different substances tried. If this happened not to be so, we might have to use slightly different values of $G$ in Eq. (5) for different pairs of particles. Presumably, there does exist a constant of gravitation $G$, the same for every pair of particles.

Equation (5) is an extension of Eq. (4), containing added information for the case of gravitational interaction [forming the ratio $\mathbf{a}_{x-y}/\mathbf{a}_{y-x}$ from Eq. (5), we get Eq. (4a)]. Hence the well-known confirmation of Eqs. (1) and (5), principally through their application in making long-range predictions of the positions of planets and satellites, is at the same time confirmation of Eqs. (1)–(4) for a gravitational interaction, but gives us no information concerning the validity of these equations in other types of interaction.

It so happens that Eqs. (1) and (4) do not apply in general, beyond a first degree of approximation, to the interaction between electrically charged particles.[10] Nor do these equations apply to the interaction between particles having velocities so high that the finite speed of light cannot be neglected in the definition of extended distance and extended time. More general equations follow.

For the case of uncharged particles having any speed up to the limiting value $c$, Eq. (2) takes the form

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathbf{v}_s(1-v_s^2/c^2)^{-\frac{1}{2}}] \Big/ \frac{\mathrm{d}}{\mathrm{d}t}[\mathbf{v}_x(1-v_x^2/c^2)^{-\frac{1}{2}}] \doteq -k_{x-s},$$

where, as before, $k_{x-s}$ is a positive numeric characteristic of particle $x$ relative to the International Kilogram; we are not concerned with electric charge in a comparison with the Kilogram. We replace the set of numerics $k$ by a primary quantity, called *rest mass* $m_0$ in the *MLT* system. By definition, the rest mass of the International Kilogram is 1 kgm, and the rest mass of particle $x$ is $k_{x-s}$ kgm. The quantity $m_0(1-v^2/c^2)^{-\frac{1}{2}}$ is called the *relativistic mass* of a particle, by definition. Thus, relativistic mass is a *secondary quantity*. I wonder if this point is generally recognized.

In m.k.s. units, Eqs. (4a) and (4b) assume the more

[8] For a brief discussion of these implications, see A. Einstein, *Meaning of relativity* (Princeton, 1945), pp. 55–56, 99–108.

[9] On p. 303, reference 1, Mach says, "Even if we adhere absolutely to the Newtonian points of view, and disregard the complications and indefinite features mentioned, which are not removed but merely concealed by the abbreviated designations 'Time' and 'Space,' it is possible to replace Newton's enunciations by much more simple, methodically better arranged, and more satisfactory propositions. Such, in our estimation, would be the following." This statement is followed by the three *experimental propositions* and the objective definitions of mass and force that were first published in 1868. Also, on p. 295, we find this statement: "There is, I think, no difference of meaning between Lange and myself . . . about the fact that, at the present time, the heaven of fixed stars is the only practically usable system of reference, and about the method of obtaining a new system of reference by gradual corrections."

[10] See L. Page and N. I. Adams, Jr., *Am. J. Physics* **13**, 141 (1945); A. J. O'Leary, *Am. J. Physics* **14**, 63 (1946).

general form

$$\frac{\mathrm{d}}{\mathrm{d}t}[m_x\mathbf{v}_x+\tfrac{1}{2}e_xe_y\mu_0(\mathbf{v}_x/r+\mathbf{v}_x\cdot\mathbf{r}_{x-y}\mathbf{r}_{x-y}/r^3)]$$
$$=-\frac{\mathrm{d}}{\mathrm{d}t}[m_y\mathbf{v}_y+\tfrac{1}{2}e_xe_y\mu_0(\mathbf{v}_y/r+\mathbf{v}_y\cdot\mathbf{r}_{x-y}\mathbf{r}_{x-y}/r^3)],$$

where $e$ denotes electric charge, $m$ relativistic mass, and $\mu_0$ permeability of space. In this form of the law of inertia, electromagnetic inertia is recognized along with inertia of matter.

*Application of Eqs. (1) and (4) to the weighing process on a spring balance.*—To simplify the discussion, let the effective acceleration of all bodies under consideration be zero with respect to the earth as the frame of reference. In its interactions with all the particles of the earth, a body $y$ on a spring balance has a net individual acceleration $\mathbf{g}$ directed downward, and all the while in its interaction with the balance $b$, it has an individual upward acceleration $\mathbf{a}_{y-b}$ equal to $-\mathbf{g}$ [from Eq. (1), $\mathbf{a}_{y-b}+\mathbf{g}=0$]. Applying Eq. (4a) to the interaction between $y$ and $b$, we get

$$m_b'\mathbf{a}_{b-y}=-m_y'\mathbf{a}_{y-b}=m_y'\mathbf{g};$$

similarly, for a different body $z$ on the same balance, we have

$$m_b'\mathbf{a}_{b-z}=m_z'\mathbf{g}.$$

Dividing the first of these equations by the second, we obtain

$$\mathbf{a}_{b-y}/\mathbf{a}_{b-z}=m_y'/m_z'. \tag{6}$$

When body $y$ is added to it, the balance acquires an individual downward acceleration $\mathbf{a}_{b-y}$ in addition to the accelerations it had before the addition of $y$; and simultaneously, the individual upward acceleration of the balance in its interaction with its support is increased in magnitude by the amount $a_{b-y}$. Acquisition of these two oppositely directed components of acceleration, of magnitude $a_{b-y}$, is indicated by a stretch in the spring of the balance; similarly, for a different body $z$ suspended on the balance. Assuming that the elongation of the spring is independent of the nature of the body suspended on it, we take the magnitude $a_{b-y}$ in one instance to be equal to the magnitude $a_{b-z}$ in another instance if the spring is stretched to the same extent in the two instances. Inserting this equality in Eq. (6) for two bodies $y$ and $z$ that produce equal elongations when suspended in turn on the same spring balance at the same point, we get $m_y'=m_z'$. If a third body $s$ stretches a spring balance to the same extent as two bodies $y$ and $z$ together, we get $m_s'=m_y'+m_z'$ as long as any difference in position of the bodies can be neglected in the measurement, so that we can take $\mathbf{g}$ to be the same for each; similarly, for any number of bodies together on a balance. It is evident that a spring balance can be used in the selection of secondary standards having values of $m'$, in kilograms, equal to 1, 0.5, 0.2, 0.1, $\cdots$. It follows that $m'$ for various bodies can be measured on a spring balance calibrated at the point where it is used with the aid of a graduated set of standards.

Taking it to be an experimental fact that $\mathbf{g}$ is the same for all particles at a given point, Eqs. (1)–(4) are confirmed by the complete self-consistency of measurements on a fine spring balance, and likewise by the self-consistency of measurements on the analytic beam balance.

### *WLT* System

*Further definitions and laws.*—With the purpose in mind of obtaining a convenient relation between local weight and standard weight in a later step, we employ the international standard magnitude of acceleration $g_s=9.80665$ m sec$^{-2}$ $\equiv 32.1740$ ft sec$^{-2}$; we may note that $g_s$ is a secondary magnitude defined in terms of the primary magnitudes for length and time. By definition, the vector quantity $w_x\mathbf{v}_x/g_s$ is called the ***linear momentum*** of particle $x$. Letting $\mathbf{p}$ denote linear momentum, we have $\mathbf{p}_x=w_x\mathbf{v}_x/g_s$. The units of momentum are 1 kgwt sec, 1 gwt sec, 1 lbwt sec, and so forth. Dividing both sides of Eq. (4b) by $g_s$, with $w$ replacing $m'$, and combining this equation with Eq. (1b), we get the law of conservation of momentum for an isolated system of $n$ particles, a corollary of the Mach law. Or, omitting the extraneous factor $g_s$ and calling $w\mathbf{v}$ the ***linmentum***[11] of a particle, we

[11] This invention of the name linmentum raises a fine point of distinction concerning the meaning to be attached to the technical term *physical quantity* in its application to two quantitative measures that are essentially the same, differing only by a constant factor other than a numeric so that their units have different dimensions. In my opinion, it is best to take two such measures to be different physical quantities, in which case the concept of dimensions applies to physical quantities themselves as well as to their units. Linear momentum and linmentum are different physical quantities in this sense of the term; their dimensions are [WT] and [WLT$^{-1}$], respectively.

get a more convenient law of "conservation of linmentum."

We may distinguish two types of force corresponding to the two types of acceleration. By definition, the quantity $w_x\mathbf{a}_{x-y}/g_s$ is called *the individual force* $\mathbf{F}_{x-y}$ *exerted on particle x by particle y*, and $w_y\mathbf{a}_{y-x}/g_s$, *the individual force* $\mathbf{F}_{y-x}$ *exerted on particle y by particle x*. Or, what is the same thing, the individual force $\mathbf{F}_{x-y}$ is the individual rate of change of momentum of particle $x$ accompanying an interaction with particle $y$; similarly for $\mathbf{F}_{y-x}$. Equations (4a) and (4b) may now be written in the form of Newton's third law,

$$\mathbf{F}_{x-y} = -\mathbf{F}_{y-x}. \tag{7}$$

The effective time-rate of change of momentum of a particle, or its effective acceleration multiplied by $w/g_s$, is called the *effective force* on the particle, by definition. The relation between these two types of force is obtained from Eq. (1). Multiplying both sides of Eq. (1a) by $w_x/g_s$, we get

$$\mathbf{F}_x = \sum_{y=1}^{n} \mathbf{F}_{x-y}, \quad (y \neq x) \tag{8}$$

where $\mathbf{F}_x$ is the effective force on particle $x$ in an isolated group of $n$ particles. The two complementary sets of relations, Eqs. (7) and (8), are merely alternative statements of Eqs. (1) and (4), the Mach law of inertia. We may say that the Mach law consists of two complementary "demi-laws" that assume meaning in a strict sense only in combination, but which may, for all practical purposes, be treated as actual laws in their own right. For convenience in applications, we may refer to any one of Eqs. (1a), (1b) and (8), as the "law of superposition." In words, Eq. (8) states that the effective force on a particle is equal to the *resultant*, or *vector sum*, of all the individual forces exerted on it in its interactions with other particles. The force method of solving problems not involving rotation is indicated in this statement. The method amounts to this: we isolate a particular particle for consideration and evaluate the effective force on it, or record the expression for the effective force in case the effective acceleration is not known; we then identify the various interactions to which the particle is subject and indicate the individual forces on it in a free-body diagram, obtain their resultant and set the resultant equal to the effective force on the particle. Statics enters as a special case of kinetics. A particle is said to be in equilibrium if its effective acceleration is zero, or, what amounts to the same thing, if the effective force on it is zero; in accordance with the law of superposition, the resultant of the individual forces on a particle in equilibrium is set equal to zero.

The law of gravitation between two particles is usually expressed in terms of force. It is most conveniently applied when written in the form

$$\mathbf{F}_{x-y} = G'w_xw_y\mathbf{r}_{y-x}/r_{y-x}^3,$$

where

$$G' = G/g_s = (6.805 \pm 0.003) \times 10^{-12}\ \text{m}^2\ \text{kgwt.}^{-1}$$

The net individual force $w_x\mathbf{g}/g_s$, peculiar to that frame of reference consisting of the particles of the earth, is called the *local weight* of particle $x$ at the particular point where the particle is located. Letting $\mathbf{W}$ denote local weight, we have $\mathbf{W}_x = w_x\mathbf{g}/g_s$. Use of the physical quantity *force* provides a somewhat simplified mode of expression in analyzing the weighing process on a spring balance but does not alter the basic argument of the preceding section. Using different words, we obtain the result that two bodies $y$ and $z$ have equal local weights, and therefore equal standard weights, if they produce equal elongations when suspended in turn on the same spring balance $b$ at the same point. Because this conclusion is so familiar, we are apt to forget that it is arrived at only as a result of five separate steps as indicated in the following chain equation: $w_y\mathbf{g}/g_s = -\mathbf{F}_{y-b}$ [Eq. (8), $\mathbf{F}_y$ being zero] $= \mathbf{F}_{b-y}$ [Eq. (7)] $= \mathbf{F}_{b-z}$ [assuming the elongation of the spring to be independent of the nature of the body suspended on it] $= -\mathbf{F}_{z-b}$ [Eq. (7)] $= w_z\mathbf{g}/g_s$ [Eq. (8), $\mathbf{F}_z$ being zero]. A similar deduction involving five separate steps may be made for the beam balance.[12]

[12] The term "massing" has been proposed by A. G. Worthing [*Am. J. Physics* **12**, 373 (1944)] as the only appropriate description of a measurement on a beam balance. Actually, it is purely academic whether one takes mass or standard weight to be the quantity measured on a beam balance or on a spring balance as the latter is used in accurate weighing, the calibration being checked with standards at the point where the balance is used. The term "weighing," meaning the measurement of standard weight, is not altogether inappropriate although it does have two disadvantages: it discriminates against mass, and it leaves room for confusion between standard weight and local weight.

The method of development has been indicated in sufficient detail so that the remaining steps need only be mentioned. Setting up the law of conservation of energy for conservative systems is largely a matter of definition. The law of conservation of angular momentum for an isolated system of particles, and the torque law, can be derived without difficulty. The Hamilton principle and other equivalent expressions can be established entirely by deduction from Eqs. (1), (4) and (5) with the addition of appropriate definitions and mathematical procedures. In brief, all the laws of dynamics can be deduced from the foundation of the preceding section with the understanding that elastic bodies, dissipative systems, rigid bodies, fluids, and so forth, are concepts introduced by definition. Or, if one prefers, the properties of such systems may be introduced as further experimental information. From the dynamics of particles alone, one does not expect to arrive at laws that depend upon the internal structure of matter, such as the Hooke law.

*Résumé.*—Because of the particular way in which force is defined in the *WLT* system, the units of the primary quantity become units of force. Or, to put it another way, the standard weight of each particle becomes a particular force magnitude; it is the magnitude of the force on the particle when it has acceleration of magnitude $g_s$; and it is the magnitude of the local weight of the particle at any point where $g=g_s$, the purpose for which the standard magnitude $g_s$ was introduced in the first place. The units of standard weight may be said to be *gravitational units* only in the sense that they are made so as an afterthought, and if they are referred to in this way, the term *absolute* should be included as a prefix. Other systems of dynamics might easily be constructed in which the units of momentum or the units of energy are made the same as the units of the primary quantity.[13]

One might choose to give the secondary quantity $w_x/g_s$ a special name, calling it the *mass* of particle $x$, by definition. However, we seldom have occasion to calculate a value of $w/g_s$ by itself, so there is not much need of a special name for this quantity. To avoid confusion, it would seem better not to use the term mass at all in the *WLT* system. In any case, the two systems of dynamics are made mutually commensurable by taking $[M]=[WL^{-1}T^2]$, $[W]=[MLT^{-2}]$, with the following relations between units, using the approximation 9.8 in place of 9.80665: 1 kgwt $=9.8$ newtons $=9.8\times10^5$ dynes; 1 m kgwt $=9.8$ j $=9.8\times10^7$ ergs; 1 m kgwt sec$^{-1}=9.8$ w.

Reflecting upon quantitative measures in physics, we may see that the items of first importance are four primary standard magnitudes; one of these is the *quantity of matter*[5] in the International Kilogram, independently of any particular value assigned to this magnitude. The significance of dimensions lies wholly in the relative dimensions of one unit with respect to other units, or, according to the meaning of physical quantity in footnote 11, the relative dimensions of one physical quantity with respect to other physical quantities. The situation concerning mass and standard weight may be summed up as follows. In anticipation of two slightly different sets of definitions for secondary quantities, the primary quantity is assigned a different name, different units and a different dimension in the two systems, even though mass and standard weight are defined and measured in identical fashion. This difference between mass and standard weight is a purely academic one created to make the respective dimensions of momentum, force, kinetic energy and the other secondary quantities equivalent in the two systems, thereby permitting us to use the same verbal language for the most part in both systems. To illustrate the latter statement, let us see what the situation would be if the primary quantity were called mass in both systems, all other procedures remaining the same. In such case, to avoid confusion between the two systems, we would want to use some name other

[13] The following example may be of interest. In this case, let us call the primary quantity *magmentum*. By definition, the magmentum of the International Kilogram is 1 newton sec[$\equiv10^5$ dyne sec] and the magmentum of particle $x$ is $k_{x-s}$ newton sec. Letting $m''$ denote magmentum, we have $m_x''=k_{x-s}m_s''$. If a standard magnitude of speed $v^s=1$ m sec$^{-1}$ is introduced, the effective momentum $\mathbf{p}_x$ of particle $x$ and the effective force $\mathbf{F}_x$ on particle $x$ may be defined by the equations $\mathbf{p}_x=m_x''\mathbf{v}_x/v^s$, $\mathbf{F}_x=\dot{\mathbf{p}}_x$. The law of gravitation may be written in the form

$$\mathbf{F}_{x-y}=G''m_x''m_y''\mathbf{r}_{y-x}/r_{y-x}^3,$$

where $G''=(6.673\pm0.003)\times10^{-11}$ m$^2$ sec$^{-2}$ newton$^{-1}$. In the *MLT* system, where mass is used in place of magmentum, the constant of gravitation $G$ has the same numerical value as $G''$ but different units and dimensions.

than momentum for the quantity $m\mathbf{v}/g_s$, some name other than force for the quantity $m\mathbf{a}/g_s$, and so forth; names of secondary quantities and verbal statements of definitions and laws would be different throughout the two systems. In this illustration, it is taken for granted that $g_s$ remains an acceleration magnitude; if $g_s$ were introduced as a numeric, we would not be defining the engineering system of units.

### Another Possible Procedure in Defining Secondary Quantities

In contrast to the method of this paper, several writers[14] have recommended that a transformation factor for converting units be included in each physical equation, the purpose being to express both systems of dynamics in the one set of equations; similarly for the different systems of electrodynamics. Let us see how this method operates in one of the initial steps in dynamics. We begin by defining a primary quantity as before; we may call it mass. Letting $K$ denote a dimensionless transformation factor, the quantity $Km\mathbf{a}$ is called force, by definition. We have $\mathbf{F}=Km\mathbf{a}$ and $\mathbf{W}=Km\mathbf{g}$. The *mks* unit of force is defined by setting $K=1$ newton/kgm m sec$^{-2}$, where the newton is a special name for 1 kgm m sec$^{-2}$; similarly for the dyne and the poundal. Metric engineering units of force are defined by setting $K=1$ kgwt/9.80665 newtons$\equiv(1/g_s)$ kgwt/kgm; similarly for British units. Although this procedure is legitimate, one can understand why so many engineers prefer an independent *WLT* system in which the kilogram mass and pound mass have no place. Simultaneous use of mass units along with weight units of the same name is an unnecessary source of confusion.

The method has other disadvantages which, it seems to me, make it undesirable even in electrodynamics, where no special complications arise from engineering practice. (i) We lose to a large extent the forthright simplicity with which each system can be developed by itself. (ii) Definitions and two different types of law all have the same form of equation and are too easily confused with one another. The two types of law may be noted in the preceding sections. We have one type of law relating physical quantities that have all been defined independently of the law itself. Each such law contains a proportionality factor except when the factor is eliminated by forming dimensionless ratios. Examples are the law of gravitation and the Hooke law; an example outside dynamics is Ohm's law, the proportionality factor being called resistance. We have a second type of law in which a proportionality factor does not appear; in this category, we have the Mach law of inertia and all other laws derived therefrom—torque law, conservation of angular momentum, and so forth. We may note further that the method of defining secondary quantities in the preceding section amounts to nothing more than the assignment of special names to particular combinations of primary quantities; a proportionality factor is nowhere involved in the definitions. In electrodynamics, we have a similar situation with respect to laws and definitions in each individual system of units.

[14] D. Ross, *Am. J. Physics* **13**, 121 (1945); A. G. Worthing, *Am. J. Physics* **14**, 137 (1946). Also, the method recommended by these writers is followed to a limited extent in some textbooks.

---

*Spiritual power is a force which history clearly teaches has been the greatest force in the development of men. Yet we have been merely playing with it and have never really studied it as we have the physical forces. Some day people will learn that material things do not bring happiness, and are of little use in making people creative and powerful. Then the scientists of the world will turn their laboratories over to the study of spiritual forces which have hardly been scratched.*—Charles P. Steinmetz.

## General Semantics and the Science Teacher

Wendell Johnson
*Psychological and Speech Clinic, State University of Iowa, Iowa City, Iowa*

JUST what science teachers teach is not always entirely apparent. What they could teach is something else again, something worth talking about, something that is doubtless worth teaching: they could teach science as a general problem-solving method, as a basic life orientation. The possibility is exciting. Just what would it involve?

It would seem to involve, first of all, a reasonably clear grasp by the teacher of the fact that the word *science* can be used in various ways, and of the various ways in which the word might be used. One of the more intriguing aspects of faculty club room bull sessions is to be glimpsed in the relative enslavement of physicists, chemists and other "scientists"—to say nothing of historians, philosophers and other "nonscientists"—by their accustomed verbal definitions. Among the eight semantic wonders of the world is to be included the apparently interminable debate over what science *is*. The furors created among the six blind men by the elephant, and among modern critics by Picasso and the late Gertrude Stein, hardly rival in turbulence the verbal whirlwinds generated among physicists, for example, by the assertion that psychology *is* scientific, and among psychologists by the assertion that it *is* not. The one question that seems effective to some degree in stilling such verbal typhoons momentarily—on occasion—is this: Precisely how does one deal with a problem nonscientifically? The difficulties we have in defining *science* are as nothing compared with the slashing about we indulge in when trying to specify how not to be scientific. In fact, it would appear that almost the only examples of quite thoroughly nonscientific behavior are to be found in the more bizarre case histories in textbooks of psychiatry and abnormal psychology.

It is difficult to believe that the need, indicated by such terminological wrestling, for even an elementary acquaintance with semantics is as widespread as it undeniably seems to be. "The uncritical assumption of mutual understanding," pointed out by Professor Quine of Harvard, is apparently far more pervasive, and produces more devastating confusion, than can easily be appreciated. The degree to which each one of us takes for granted that everyone else is familiar with the nooks and crannies of his own "lonely room" of words and meanings is a formidable barrier to communication. It is this semantic provincialism and individualism that underlies much of the seemingly unaccountable disagreement and general futility of formal and informal discussion. We waste our time and unravel our nerves contending that the speaker "isn't defining his terms correctly" and "can't speak English"—"correctly" and "English" referring to our own pet, and unavoidably private, definitions and ways of using words.

One of the more elementary questions in this connection, a question important to science teachers, has to do with the purposes served by definition: What is a definition for? There are several possible answers. Perhaps the one that would be given most commonly is that a definition specifies what a word refers to, and so serves the purposes of understanding and agreement. Doubtless, there are few who would not accept this statement, at least in principle—and if permitted the qualification that very often definitions do not seem in practice to serve the indicated purposes. In any event, however, the statement hardly exhausts the subject. It is too large a subject to be exhausted, here at least, but there is one other answer to be given to the question that will be useful to us in the present discussion: A particularly important function of a definition is to be seen in the effects that it has on those who use it. These effects can be limiting, inhibitive and restrictive. On the other hand, they can be broadening, facilitative and stimulating. *Music*, for example, can be defined in such a way as to deaden one's capacities for enjoying anything but Beethoven—or Spike Jones. Or it can be so defined as to stimulate

one's curiosity and exploratory drives and prepare one to enjoy *both* Beethoven *and* Spike Jones.

As with the word *music*, so with the word *science*. It can be defined in such a way as to narrow one's interests, restrict one's activities, and limit the usefulness and value of one's "science." One can, and many people do, operate with a definition of *science* such that it never occurs to them that problems which they classify as "moral," or "ethical," or "religious," or "personal" could be dealt with scientifically. The term can be defined, however, in ways that encourage a scientific approach to an ever-widening range of situations. If teachers of physics believe that science provides an effective problem-solving method, one might hope that they would react with some enthusiasm to definitions of *science* that would extend the usefulness of the word as much as possible. Such definitions provide illuminating slits in the iron curtain of traditional professional exclusiveness and rigid departmentalism, slits through which one can glimpse the possible scientific civilization of the future.

For practical purposes, one can distinguish at least five general meanings of *science*. The term is frequently used, particularly in newsstand magazines and in movies, to refer to certain technics, usually involving the use of specialized apparatus. The more Rube Goldbergish the apparatus, the more "scientific" it is, supposedly. The men who operate the apparatus and apply the technics are said to be scientists. Almost always, however, such men might more usefully be designated as technicians. Many a good technician would be thoroughly immobilized if asked to give a lecture on the basic principles of scientific method. As a means of making a useful distinction I have proposed the term *tinkerniquer*, in order to differentiate the laboratory workers who "tinker" with technics from the scientists who plan their labors and interpret their findings. The distinction is by no means intended to be derogatory. It is intended merely to facilitate a useful differentiation. The distinction is pointed up somewhat by the fact that Einstein has spent little, if any, of his working time with apparatus.

*Science* is also used sometimes to refer to what are called "the sciences." It serves, thus, as a kind of verbal wedge, splitting off physics, chemistry and a few other academic areas, from all other areas. A considerable proportion of word-workers devote themselves to driving this wedge here and there, according to their acquired preferences, into the tree of knowledge. Used as a wedge in this way, the word *science* tends to make for statements, monographs and books that discourage attempts to behave scientifically outside certain lecture halls and laboratories. They discourage communication and comradeship as between those who work in "the sciences" and those who toil in "the humanities" or "the arts." The effect of this verbalistic splitting is individually and culturally schizophrenic. The effect is seen in the physicist who steps out of his laboratory and into the sixteenth century each afternoon at five o'clock. It is seen no less in the historian who, convinced that he is in no sense a scientist, abandons himself with various and sometimes curious effects to what he calls the art of storytelling.

*Science* is used in still another way as referring to "scientific knowledge." This implies that there can be, and actually is, "nonscientific *knowledge*." One must wonder what the purveyors of "scientific facts" would recognize as "unscientific *facts*." If "unscientific facts" is a contradiction in terms, then "scientific facts" is a redundancy. Outside this semantic eddy, the definition proceeds with its cargo, supplying the writers of popular magazine pieces, the ads, the believe-it-or-nots, and high school and college textbooks with a never-ending assortment of more or less reliable items of information about everything from fossils and fungi to tea and tumors. From this point of view scientists are essentially hucksters, dispensing oddments of knowledge in sets or by the piece. This quiz-show conception of science is hardly calculated to give the students who use the textbooks or the laymen who read the digests any very workable notions as to where the "facts" come from, the methods of obtaining and validating—and revising—them, or the broad significance of their implications as they ripen and expand in the yeast of theory. It emphasizes the results of science at the expense of the method, and without providing for systematic evaluation of the results themselves.

There is also the definition of *science* that focuses attention on theory, as such. Science, as viewed from this angle, is a body of hypotheses, theories and laws. Well-developed scientific theories necessarily function on relatively high levels of abstraction, and effective understanding of them presupposes a familiarity with mathematics, with the principles of theory construction, and with the data basic to any specific theory being considered. Few people are qualified by training or experience to evaluate such theories, to say nothing of formulating them, and if others are given to understand that these theories are what *science* mainly refers to, their interest in it can be expected to be slight at best. This is not to say that the majority of people are inherently incapable of appreciating the theoretical aspects of science; it is only to say that our customary schooling, even our science courses, seem not to provide the necessary training.

There is, however, a definition of *science* that, in contrast to those sketched above, tends strongly to emphasize the generalization of science into practically all areas of human endeavor. This is the definition that emphasizes the methodological aspects of science, presenting it as a problem-solving method. The basic method that we call scientific would appear to be essentially the same in all the so-called sciences—and outside the recognized sciences—and is to be clearly differentiated from specific technologies and narrowly applied technics. Even precision apparatus, rigorous quantification and experimental controls are not *essential* to scientific method; they are *refinements* of it. The basic method can be applied to any sort of problem whatever, inside or outside the laboratory. In this sense there can be such a thing as scientific living. It is this basic method that science teachers *could* teach—and it is invigorating to contemplate the possible consequences of teaching it well to the succeeding waves of incoming students that threaten perennially to make the dry academic beaches damp with promise.

Science, from this point of view, can best be regarded as a general orientation, to be contrasted with what might be called a prescientific orientation. This contrast may be expressed in the following broad terms:[1]

## BASIC FEATURES OF PRESCIENTIFIC ORIENTATION

*1.* Fundamental notion of the static character of reality. A static reality involves essential constancy (there is nothing new under the sun). Main attention is given to similarities; differences are minimized or ignored. Consequently, the individual is not especially important except as he represents a type.

*2.* Rigidity, or conservatism, the tendency to maintain established beliefs and habits regardless of changing conditions, is fostered by these basic notions of static constancies. Thus, traditions are cherished, and the authority of age and precedence is extolled, seldom challenged; experimentation is discouraged. The Old Man is honored and obeyed. As a result of all this, individual infantilism and social retardation are fostered.

*3.* The basic method of problem-solving, which we call authoritarian, involves mainly the practice of abiding by advice obtained from some vested authority, such as a parent, teacher, priest, or judge. Authority sometimes resides also in a book or code of rules. The pronouncements of such authority are not to be revised. This authoritarian method works in practice to maintain unchanged the traditional beliefs, customs, and rules of conduct. If prob-

## BASIC FEATURES OF SCIENTIFIC ORIENTATION

*1.* Fundamental notion of the process character of reality. A process reality gives rise to a never-ending series of differences. As much or more attention is paid, therefore, to differences as to similarities. As one important consequence, the individual is regarded as an individual, not merely as an example of a type.

*2.* Adaptability, a readiness to change as changing conditions require, is fostered by these basic notions of process differences. Thus there is a tendency to challenge authority systematically; to experiment, to test traditional beliefs and customs against actual observation and experience. The Old Man is respected but evaluated critically. As a result of all this, individual and social maturity is stimulated.

*3.* The basic method of problem-solving, which we call scientific, consists of four main steps: (*a*) the asking of questions that direct one's (*b*) observations so as to (*c*) answer the questions clearly in such a way as to test one's beliefs or assumptions, (*d*) which are revised accordingly. Of these four steps, three (*a*, *c* and *d*) involve mainly the use of language. This scientific method works in practice toward the continual improvement of specific technics, re-

[1] For a more extended statement see W. Johnson, *People in quandaries: the semantics of personal adjustment* (Harper, 1946).

lems are not solved, they are "explained" in terms of "fate," or "nature," or the "supernatural"; and toward the *language* used in such "explanations" there is a dominant attitude that is naïve and unreflective.

*4.* The language of a prescientific orientation is designed to control behavior by virtue of the vested authority it represents. If it is not clear, a properly appointed authority will interpret it, and his interpretation is to be believed. The validity of authoritarian pronouncements is not to be questioned. Statements of assumption and statements of fact tend to be regarded as the same.

*5.* Prescientific language tends to make for questions that are frequently vague and quite often meaningless factually. The observations needed to answer the questions, and the procedures needed to make the observations and to indicate their degree of reliability, are neither specified nor clearly implied. Attempts to answer such questions give rise to misunderstandings and disagreements, to misinformation and misleading theories, with the result that predictability and foresight are achieved slowly, or not at all, and individual and social maladjustments are thereby fostered.

*6.* It is realized only vaguely, or not at all, that every statement conveys information about the speaker as well as information about whatever the speaker may seem to be talking about; and the degree of self-reference is largely ignored in evaluating the statement's factual significance.

*7.* In a prescientific orientation, there is a marked tendency to speak as though with the voice of another (ventriloquizing). For example, the voice of The Law is not recognized as the voice of the Judge himself. The speaker tends to ventriloquize both unconsciously and deliberately (as in the planned use of "ethical proof"). Only the more artful and deliberate ventriloquizers seem to realize that, after all, it is their own evaluations that they are expressing. Students are encouraged to take for granted the validity of statements made by their teachers and by the authors of their textbooks.

*8.* Accurate prediction, or foresight, is not a particularly well-recognized objective in a prescientific orientation. At least, theories and specific statements are not evaluated primarily in terms of their usefulness in making predictions. In a prescientific orientation there are, strictly speaking, no scientific submicroscopic theories; there are, rather *beliefs* regarding the "supernatural," for example. These tend not to be changed, because they are considered not as theories but as statements of fact. Faith in these beliefs and obedience to the authority that represents them—obedience expressed by participation in prescribed rituals, for example—are prized as the means of control over natural and human events.

finement of beliefs, and "modernization" of customs and rules of conduct. If problems are not solved, new theories and methods are devised to solve them.

*4.* The language of a scientific orientation is designed to be factually meaningful, directly or indirectly, and clear and valid. It is intended to satisfy two important tests: "What do you mean?" and "How do you know?" Moreover, assumptions are sharply differentiated from statements of fact.

*5.* Scientific language is oriented around factually clear, answerable questions. Terms are defined operationally by reference to relevant investigative procedures. Vague or meaningless questions are abandoned as being misdirective of human energy. On the principle that the terminology of the question determines the terminology of the answer, only clearly stated questions are tolerated. Because of this, mutual understanding and agreement are facilitated, predictability and foresight are improved steadily, and individual and social adjustment are thereby fostered.

*6.* It is realized that every statement conveys information about the speaker as well as information about whatever the speaker may seem to be talking about; and the degree of self-reference is reckoned in evaluating the statement's factual significance.

*7.* In a scientific orientation, there is little or no tendency to speak as though with the voice of another (ventriloquizing). For example, the voice of The Law is recognized as the voice of the Judge himself. The speaker tends not to ventriloquize either unconsciously or deliberately; he realizes that what he expresses are his own evaluations—even though he may quote another man's words. Students are encouraged, and are shown how, to check the validity of statements made by their teachers and by the authors of their textbooks.

*8.* Accurate prediction, or foresight, is a clearly recognized objective in a scientific orientation. Theories and specific statements are evaluated primarily in terms of their usefulness in making predictions. The value of a scientific submicroscopic theory (such as a molecular theory of matter) lies in the accuracy of the predictions which it makes possible. Changes in such theories, as also in theories that do not clearly involve submicroscopic constructs, are made in the interests of more adequate prediction. Theories of high predictive value are prized as the means of control over natural and human events.

Viewed in this way, as contrasted with a prescientific approach to situations and problems of whatever kind, science comes alive abundantly with personal and social implications. It overflows the laboratory dikes and floods the entire academic landscape with waters of unsuspected fertility. Nor does it dry up at the boundaries of the campus. Science, as defined here, provides ways of approaching political, international, ethical, moral, industrial and personal problems, as well as technological ones. As Professor S. A. Nock[2] has said, it is not what he works with but

[2] S. A. Nock, "The scientist and ethics," *Ethics* **54**, 14–28 (1943–44).

the method by which he works with it that distinguishes the scientist.

One may, of course, insist that this is not the definition of *science* to which one is accustomed—but such a statement would be appropriately followed, it might reasonably be expected, by a defense of its relevance to some significant proposition. Otherwise, the statement would be properly regarded as autobiographical, or self-assertive, and one would find it more or less interesting, depending on one's personal relationship to its author. In the meantime, it is to be considered that the definition advanced above has, like any other definition, an effect on those who operate with it. The effect in this case might be expected to have the advantages, so far as problem-solving is concerned, that we have come to associate with the use of scientific method in those areas where it has so far been applied.

It is not to be assumed that most teachers of physics—or of chemistry, for example—have either the qualifications or the time to train their students in the specific details of the generalized use of scientific method. It is to be assumed, however, that they are, or should be, very well qualified—and that they do have the time and the opportunity—to present science as a method which can be applied far beyond the boundaries of their own restricted fields. It is reasonable, moreover, to expect them to make the most of this opportunity. Another generation of technicians and engineers, with technological reflexes, incompetent to assist in dealing with the social problems which their technical competence creates, is indeed a terrifying and demoralizing prospect. The science teacher who views it with schizophrenic indifference is the product of a kind of education that should, in the interests of public welfare, be abandoned. Most of the research workers who served as midwives to plutonium were sufficiently shocked when they saw the baby to offer their services as wardens and guards. Some of them even volunteered to drown it. Those of us who still think it is a cute baby might well begin our re-education by studying ourselves in a mirror for three hours a day, if we can stand it, until significant changes appear. If it does not take too long there may still be time.

So long as we continue to change the world we live in by applying scientific method to our material problems, we can hardly anticipate anything but increasing personal and social maladjustment if we continue to cherish and preserve our traditional prescientific ways of living in that world. In space-time terms—that is to say, in terms of the time it takes to get from one place to another, or to deliver a rocket bomb from one nation to another, or to send a message around the earth—we have reduced the former expanse of the globe to the size of a small state, scarcely larger than Rhode Island used to be. Into this tiny and steadily shrinking space we are packing the age-old hatreds and conflicting loyalties that once, like the submicroscopic particles of the uranium atom, were comparatively harmless because they were so far apart. But just as we released the fury of the uranium atom by bringing enough of them close together fast enough, so we are threatening to unleash the demolishing fury of the "culture atom" by compressing the peoples of the earth into compact space-time proximity so rapidly that they are unable fast enough to neutralize their accustomed prides and prejudices.

The required neutralization process has been facilitated so slightly and slowly, if at all, by our traditional prescientific institutions and methods of social control and individual development, that only an irresponsible optimist would continue to place his trust in them. A generalized scientific orientation, imparted intensively to the young—and to the middle-aged and old if possible—offers at least a promising alternative, a gambler's chance, perhaps. The scientific world we have created might, with good fortune, be held together by people who are as scientific about themselves and their social relationships as they are about their industrial and military endeavors. But people who are, by cultural tradition and intimate family and community indoctrination, predominately prescientific can hardly be expected to develop a scientific orientation to personal and social problems unless they are given the necessary training. This means there must be teachers who can train them. If science teachers can function as teachers of science as here defined, there is some reason for hope. Certainly, little assistance—and probably considerable frustration of our efforts—is to be

looked for from teachers in those areas where prescientific ways and points of view are fostered and fiercely defended.

Those science teachers who find in these remarks something to which they can respond with enthusiasm can find, also, in general semantics a reasonably systematic and comprehensive rationale and practical know-how to assist them in translating their enthusiasm into action.[3]

Anyone significantly interested in knowing about general semantics will not be satisfied by a brief definition of it. A short statement, however, might serve a useful purpose by indicating at least roughly the essential character of the problems with which general semanticists are particularly concerned. They are concerned, first of all, with the sort of problems that have been discussed in this paper, as pointed up, for example, in the contrasting characterizations of scientific and prescientific orientation presented above. General semantics amounts to a generalized and practical formulation of science as method, with due emphasis on the language of science. In teaching general semantics one teaches scientific method in such a way as to make its general value apparent. And to a considerable degree one does this by making clear the differences between the ways in which scientists use language and the ways in which the "average man"—and, by sharper contrast, the "mentally" ill person—uses language.

The word *language* is being used here in a broad sense—not merely, not primarily, in the sense of grammar or logic. General semantics is not "the study of words" or "the study of meaning," as these terms are ordinarily understood. It is more nearly correct to say that general semantics is concerned with the assumptions underlying symbol systems and with the personal and cultural effects of their use. It is concerned with the pervasive problem of the relation of language to reality, of word to fact, of theory to description and of description to data—of the observer to the observed, of the knower to the knowable. It is concerned with the role of language in relation to predictability and evaluation—and so in relation to the control of events and to personal adjustment and social integration.

As a frame of reference, an organizing system, general semantics may be described as a set of basic premises, working principles and practical technics. The major premise, that of nonidentity, summarizes the proposition that no two things or experiences are identical in all respects, and that no one thing stays the same in all respects. With certain systematic restrictions, it is to be contrasted with the basic Aristotelian premise, or law, of identity, $A$ is $A$. In comparable symbolism, the non-Aristotelian premise of non-identity may be stated as $A_1$ is not $A_2$, or $A^{1940}$ is not $A^{1930}$. The generalized formulation of non-identity is to be made in terms of the process of abstracting, a process to be described in terms of the nonsymmetrical relationships as between levels of evaluation or experience—levels of abstraction.

As originally formulated by Korzybski,[3] and adapted to their various purposes by other general semanticists, the basic distinction is that between the nonverbal and verbal levels of abstraction. The nonverbal levels are further differentiated into the macroscopic and microscopic direct experience levels, on the one hand, and the submicroscopic level of inferential data (atoms, molecules, electrons, for example), on the other. The verbal levels are differentiated into the first-order verbal level of naming and description and successively higher levels of inference. Within this structural framework, then, abstracting is viewed as a process of organizing, ordering, or relating the data of experience by leaving out or disregarding more and more experiential details as one proceeds from lower to higher levels. A consciousness of abstracting, the principal objective of training in general semantics, functions as an awareness of the details left out and those left in, and of the organizing principles (assumptions, purposes, etc.) by

[3] The original source book of general semantics is Alfred Korzybski's *Science and sanity: an introduction to non-Aristotelian systems and general semantics* (Science Press, 1933; rev. ed., 1941). Other books, which might serve as suitable introductions to general semantics and certain of its practical uses, are the following:

S. I. Hayakawa, *Language in action* (Harcourt, Brace, 1941).

W. Johnson, *People in quandaries: the semantics of personal adjustment* (Harper, 1946).

I. J. Lee, *Language habits in human affairs* (Harper, 1941).

See also *ETC.: A Review of General Semantics*, under the editorship of Dr. S. I. Hayakawa, Illinois Institute of Technology.

means of which this selective process is governed and by means of which the selected data are ordered. Abstracting, therefore, is seen as the generalized mechanism underlying evaluation. Within this framework, the premise of non-identity assumes a form that may be stated variously: $level_1$ is not $level_2$, or the word is not the object, or inference is not description, or the perceived object is not the construct or inferential event, etc.

Such matters as the self-reflexive character of language and of abstracting generally (abstracts of abstracts, statements about statements, evaluations of evaluations, etc.), the self-projection involved in observation and in verbal statement, the dependence of a word's "meaning" on the relative level of abstraction on which it is used, the self-corrective potentialities of the process of abstracting, and the problems of symbolization and communication may be meaningfully treated in terms of the process of abstracting as formulated in general semantics. These are matters of basic importance in a consideration of science as method, whether in its restricted applications or its generalized sense.

The working principles of general semantics may be summarized as those of probability (or uncertainty), which facilitates a relativistic, as opposed to an absolutistic, orientation; symbol reaction (or conditionality), which is directive of reactions governed by an awareness of the difference between a symbol and its varying non-verbal level reference (or lack of such reference), between an inference and its supporting, or refuting, descriptions, between an object or event and its relational implications, etc.; and extensionalization, which embodies and generalizes the operational principle stated by Bridgman[4] and others.

On the basis of these principles, a considerable number and variety of behavioral and evaluational technics have been devised and applied by Korzybski and others. Many of these are linguistic technics designed to make the language of every day accord more fully with the distinctive features of scientific language. Others are nonverbal modes of application of the semantic principles.[5]

These premises, principles and technics, which define the basic framework of general semantics, constitute a generalized formulation of science *as method*. It is the sort of formulation that makes scientific method teachable in a nontechnological, general education, liberal arts sense. In the condensed form in which it has here been sketched, it may possibly appear a bit formidable from the point of view of a college freshman. This is by no means the case, however, when it is expanded and explicated in a sizable book or in a three semester-hour course. It is to be hoped that no reader will be satisfied with this digest or sketch; it is meant to be only suggestive, an expository *hors d'oeuvre*. From the brief and general statements made here the science teacher may, however, be able to get at least a rough indication of what he can find in more comprehensive sources.

This article will have served its intended purpose if it arouses in the science teacher an enlarged sense of his social responsibility—and of his opportunity to meet it more adequately. When viewed broadly as a generalized method of problem-stating and problem-solving, as a kind of life orientation, science expands enormously in social significance. The science teacher should get both comfort and increased motivation from the proposition that what we need, as individuals and as a world society, is not less science but more of it. Sanctuary is not to be found by fleeing back to the Middle Ages, but by hurrying ahead into the future in order to catch up with ourselves. Our left foot is dragging, so to speak, and we are far more likely to regain our balance—or to achieve it for the first time in history—by thrusting the left foot forward than by trying to pull the right one back. Having used science to transform the world in which we live, we face the sheer necessity of using science to transform us who live in that world. And general semantics is designed to facilitate the urgent transformation.

[4] P. W. Bridgman, *The logic of modern physics* (Macmillan, 1932).

[5] For a description of several of these technics see reference 1, particularly chap. 10.

# Laboratory Equipment for a Course in Electronics

K. S. Lion
*Department of Biology, Massachusetts Institute of Technology, Cambridge 39, Massachusetts*

THIS paper describes a set of model panels for demonstrating electronic circuits in a course in instrumentation (physical measuring and research methods) which is offered to students of biology and medicine. Experience with this method seems to justify its application for students in physics and other fields as well.

The section on electronics in this course begins with a study of thermionic emission and the diode. The model panel for experiments on these topics is shown in Fig. 1*A* and, like all the others, it consists of a Masonite board $15\times 9\times \frac{1}{4}$ in. The schematic diagram of the tube with the necessary connections is painted on the front, while the corresponding wiring is on the back of the panel. The electronic tube is held in a socket in the upper left-hand corner. For laboratory work, the panel is put on a sloping support, as in Fig. 1, *A* and *C*; for lecture demonstrations, it is held vertically on a laboratory stand by means of clamps, as in Fig. 1*B*. All connections to voltage sources, resistances and meters used in the experiments are made with banana plugs and jacks as indicated.

The student, after having familiarized himself with the theoretical part of the experiment, uses this board to make the necessary connections and measurements. The schematic diagram of the experimental set-up is a substantial help to him in distinguishing clearly between the different circuits and in understanding their operation. He also gets used to grouping his set-up systematically, which enables him to build his own more complex instruments and amplifiers later.

The measurements on the triode and pentode follow as a next step. The model panels for these experiments are shown in Fig. 1, *B* and *C*, and are used in the same way as outlined for the diode. Figure 2 shows the panels for the principal types of amplifier studied in this course. The necessary coupling elements, such as resistors, capacitors and potentiometers, can be plugged in the jacks provided, as indicated in Fig. 2*F*. It may be noted that the jacks for the meters in all these models are at, or close to, ground potential. The student soon gets used to this precaution and acquires the correct measuring technic, which is essential in high-impedance or in high-frequency circuits.

Different kind of rectifier circuits are shown in Fig. 3. Student experimental work on these circuits, prior to the use of the model panels, used to be fairly difficult. The measurement of the alternating and direct potentials and currents, and of the average, rms, peak and ripple values and wave forms in the different circuits require much understanding of the physical process. This is considerably facilitated by the use of the panels. Some of the circuit elements are permanently built in, as the filament and plate transformers in the panels of Fig. 3, *A* and *B*, and the capacitors in Fig. 3, *C* and *D*. Without these elements the panel would lose the characteristic appearance of the particular circuit.

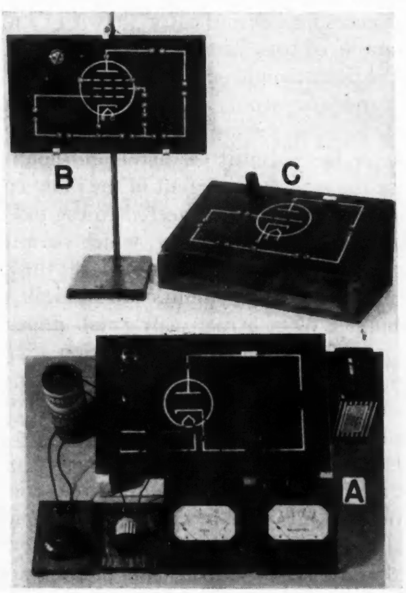



FIG. 1. *A*, model panel of the diode with meters and voltage sources connected; *B*, pentode mounted on laboratory stand for lecture demonstration; *C*, triode, mounted on cradle for use in the laboratory.

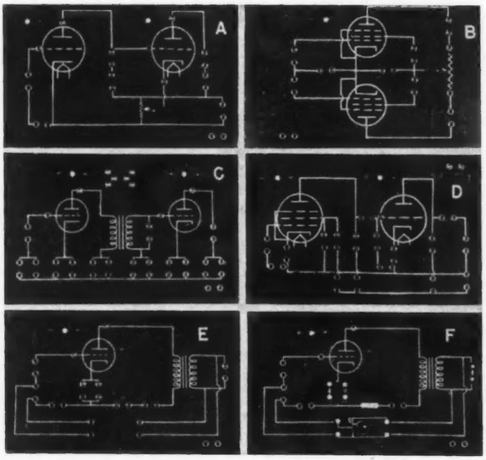



FIG. 2. Fundamental electronic circuits I, amplifiers: *A*, d.c. amplifier; *B*, balanced amplifier; *C*, transformer coupling; *D*, *RC* or *LC* coupling; *E*, feedback amplifier; *F*, same, with circuit elements plugged in.

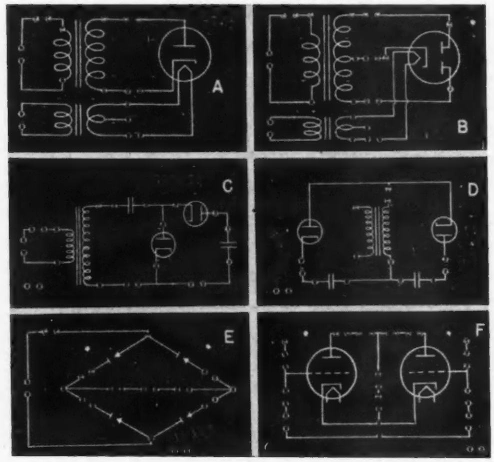



FIG. 3. Fundamental electronic circuits II: *A*, half-wave rectifier; *B*, full-wave rectifier; *C* and *D*, voltage doubler; *E*, bridge rectifier; *F*, trigger or multivibrator circuit.

It is relatively easy to group together several of these element panels and to build up a more complex instrument or to illustrate the principle of such an instrument. The potentiometer circuit with electronic null indicator shown in Fig. 4 is an example of this. The rectifier circuit together with the filter circuit (shown as a di-section filter with capacitor input) furnishes the plate potential for the vacuum tube voltmeter. This voltmeter has a sensitive meter and compensation for the quiescent current in the plate circuit. Grid and cathode are connected to the indicator terminals of the compensator, which permits the measurement of small potentials in the usual way. Most electronic circuits can be built up in this manner from a relatively small number of elements. The simplicity of operation is a particular advantage for lecture demonstrations. It is usually possible to make the necessary connections during the lecture and to develop a more or less complex circuit before the eyes of the students. Most of the elementary electric and electronic circuits used in the course are made up in a similar way; some of them are shown in Figs. 5 and 6. On the front side of the panel in Fig. 5 is a schematic diagram of a cathode-ray tube with the usual connections. A small cathode-ray tube—for example, RCA type 913 with 1-in. screen, for laboratory work, or RCA type 2AP1 with 2-in. screen, for demonstrations in a small classroom—is inserted through a hole in the "screen" of the painted diagram and connected in a socket. The result of any manipulation on the diagrammatic model can then be seen on the screen of this tube. With little difficulty the influence of any electrode potential upon in-

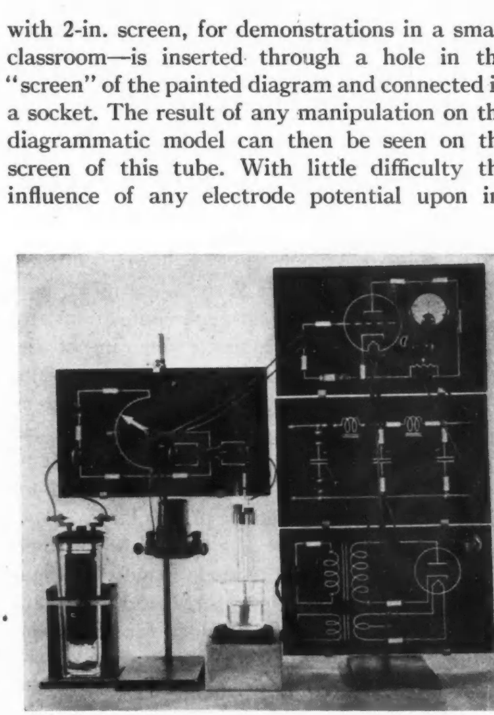


FIG. 4. Potentiometer circuit with electronic balance indicator, mounted for lecture demonstration. The indicator consists of a vacuum tube voltmeter (upper panel), power supply (lower panel) and filter (middle).

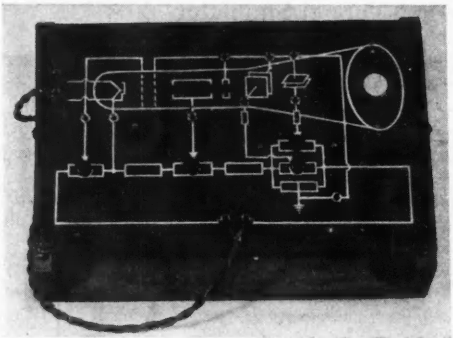


FIG. 5. Cathode-ray tube panel.

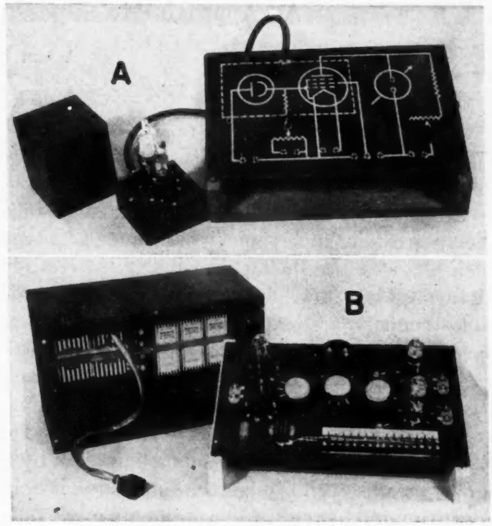



FIG. 6. *A*, photoelectric cell with amplifier; *B*, two-stage balanced d.c. amplifier.

tensity, focus and deflection can be demonstrated and measured.

Figure 6*A* shows a photo-amplifier. Such an arrangement requires careful shielding against light and electric pick-up. To accomplish this the photoelectric cell, amplifier tube and grid-leak resistor (indicated on the model panel within dotted lines) are mounted within a metal cabinet, and connected to the model panel by means of a plug and a cable. The photoelectric cell and the high-impedance grid circuit of the tube are, therefore, sufficiently shielded, while all terminals on a low-impedance level are accessible on the panel for measurements.

Shielding is equally necessary in multistage amplifiers. The plug-and-jack method can then no longer be used, and regular wiring and soldering in shielded cabinets is needed. A set-up for such work is shown in Fig. 6*B* and illustrates a two-stage balanced d.c. amplifier. All parts are mounted on a metal panel, with the exception of the fixed resistors which the student selects after computing their resistance. The necessary connections between these elements are soldered, and the amplifier panel is connected, by means of a multiple cable and plug, to the batteries in the same compartment. This kind of work is usually done at the end of the course and leads to the customary electronic technic applied in experimental or commercial instruments.

# RECENT MEETINGS

## Western Pennsylvania Section

The regular Fall meeting of the Western Pennsylvania Section of the American Association of Physics Teachers was held on November 9, 1946 at the University of Pittsburgh. Thirty-five members and guests were present. The following papers were presented.

**The mks and MKS systems of electric units.** A. G. WORTHING, *University of Pittsburgh.*

**Avogadro's law and the atomic bomb.** M. F. SERENE, *Ambridge High School.*

**Atomic charts.** J. T. SHRINER, *Allderdice High School.*

**Is one enough (in a physics definition)?** R. C. HITCHCOCK, *State Teachers College (Indiana, Pa.).*

**Demonstration of elastic impacts.** R. M. BELL, *Washington and Jefferson College.*

**My year at the Cavendish Laboratory.** R. C. COLWELL, *West Virginia University.*

**Boyle's law—tilting J-tube.** H. W. HARMON, *Grove City, Pa.*

New apparatus was informally demonstrated by O. H. Blackwood, W. N. St. Peter and R. C. Hitchcock.

At the business meeting the officers reported that the Pittsburgh high schools are planning to resume laboratory work in physics when facilities can be restored, but that this cannot be soon.[1] On motion of O. H. Blackwood it was voted to compliment the school authorities of Butler, Pennsylvania on the fact that over half of the eligible high school population of some 500 is taking physics. Dr. R. C. Colwell was elected vice president of the Section.

CHAS. WILLIAMSON,
*Secretary*

[1]See *Am. J. Physics* **14**, 341 (1946).

# An Approximate Supersonic Wind-Tunnel Simulator

Allen H. Schooley
*U. S. Naval Research Laboratory, Washington 20, District of Columbia*

The present widespread interest in devices that travel through the air with velocities greater than the velocity of sound has intensified the interest in supersonic wind tunnels. Many such devices are being constructed to study high-speed air flow.[1] However, since supersonic wind tunnels are expensive to build and to operate, they are available for only the most urgent problems; that is, many workers and students are denied the opportunity of gaining experience with such devices. Therefore it appears desirable to simulate supersonic flow by a simple means, even though the simulation is only approximately correct. Measurements made on a device fulfilling this need are described in this paper.

The similarity between the surface waves created by a boat traveling through water with a velocity greater than the velocity of propagation of the waves, and the compressional or shock waves created by an object, such as a projectile, traveling at a speed greater than the velocity of sound, is well known.[2] This is the basis for the approximate supersonic wind-tunnel simulator.

We begin by listing the most important quantities that affect the drag force $D$ tending to retard an object traveling through a gas, such as air, or on the surface of a liquid, such as water. (Table I summarizes the definitions of all terms used in this paper.)

(i) Velocity of the object, $v$. The drag tending to retard an object moving in air or floating on water is increased with increased relative velocity.

(ii) Density of the fluid, $\rho$. The drag is greater for the denser fluids.

(iii) Projected area of the object in the direction of motion, $A$. A relatively large projected area produces a relatively large drag. For floating objects the projected area is assumed to be the projected area under the surface.

(iv) Viscosity of the fluid, $\mu$. Drag is greater when the viscosity is large.

(v) Length of the object, $L$. The viscous drag force will be greater if the object is long.

(vi) Yaw, or angle of attack of object, $\alpha$. If a relatively long object is moved with its axis at an angle to the direction of motion, the drag is increased.

(vii) Velocity of wave propagation, $c$. The velocity of wave propagation affects the drag force when an object is traveling with a velocity greater than the velocity of wave propagation because, under such conditions, waves are maintained that derive their energy from the object. At velocities less than the velocity of wave propagation, no waves are formed.

We thus assume that there are a total of eight quantities, including drag, that are related to the retardation of an object traveling in air or on the surface of water. The units of these eight quantities can be expressed in terms of the three fundamental units of force, length and time. According to the Buckingham $\pi$-theorem, there are five related dimensionless groups that can be formed by combination of the quantities. If these groups are called $\pi_1, \pi_2, \pi_3, \cdots$, the relation connecting them is

$$\pi_1 = f(\pi_2, \pi_3, \cdots). \tag{1}$$

By dimensional analysis it can be determined that $\pi_1 = D/\rho v^2 A$, $\pi_2 = c/v$, $\pi_3 = \mu/\rho v L$, $\pi_4 = \alpha$, and

Table I. *Definitions of symbols.* All numerical subscripts, except in the case of $\pi$, indicate that the quantity is related to model measurements made in water.

- $a$, projected area of wing on wing chord surface.
- $A$, projected area of an object in a plane perpendicular to fluid flow.
- $c$, velocity of wave propagation.
- $C_D$, coefficient of drag.
- $C_F$, coefficient of lift, or coefficient of cross-wind force.
- $D$, drag.
- $f$, function.
- $g$, acceleration due to gravity.
- $L$, length of an object.
- $F$, lift, or cross-wind force.
- $M$, Mach number.
- $R$, Reynolds number.
- $T$, coefficient of surface tension.
- $v$, velocity of an object relative to a fluid.
- $\alpha$, yaw, or angle of attack.
- $\theta$, Mach angle.
- $\lambda$, wavelength.
- $\mu$, coefficient of fluid viscosity.
- $\pi$, dimensionless quantity.
- $\rho$, fluid density.

[1] "Faster than sound," *Science Illustrated* (May 1946).

[2] E. Preiswerk, "Applications of the methods of gas dynamics to water flows with free surface," Parts I and II, Trans. in *NACA Technical Memorandums* 934 and 936 (1940).

$\pi_5 = L/A^{\frac{1}{2}}$. Then

$$\frac{D}{\rho v^2 A} = f\left(\frac{c}{v}, \frac{\mu}{\rho v L}, \alpha, \frac{L}{\sqrt{A}}\right); \tag{2}$$

or

$$D = C_D(\tfrac{1}{2}\rho v^2 A), \tag{3}$$

where

$$\tfrac{1}{2}C_D = f\left(\frac{c}{v}, \frac{\mu}{\rho v L}, \alpha, \frac{L}{\sqrt{A}}\right). \tag{4}$$

Equation (3) is the familiar Newtonian law of fluid resistance, where $C_D$ is called the *coefficient of drag*. The reciprocal, $v/c$, of the first argument of the function in Eq. (4) is called the *Mach number*. As is well known, it is an important factor in studying the characteristics of bodies traveling with supersonic velocity in a gas. It is usually not used in the study of the characteristics of bodies traveling on the surface of a liquid. The Froude number is conventionally used as one of the important factors in studying the drag on ship models; this number is $v/(Lg)^{\frac{1}{2}}$, where $g$ is the acceleration due to gravity. It would have resulted from the dimensional analysis if $g$ had been used as one of the quantities included in the foregoing list. This was not done because the velocity of wave propagation $c$ is a function of $g$ in the case of water waves, and, in addition, the omission of $g$ makes possible the use of the Mach number technic in the simulation. Surface waves have the general nature of compressional waves in air, where a compression in air corresponds to a rise in surface level, and a rarefaction to a drop in surface level. As will be seen, there is some advantage to applying the Mach number concept in examining the effect of the formation of surface waves.

The reciprocal, $\rho v L/\mu$, of the second argument of the function in Eq. (4) is called the *Reynolds number* and is of importance chiefly under conditions where the viscous forces are relatively large. The next argument in Eq. (4) is effective where the yaw angle (ballistics), or angle of attack (aerodynamics), is not zero. The last argument of Eq. (4) is a simple form factor that comes out of the dimensional analysis and is by no means a complete specification of form.

Shock waves for a symmetrical missile traveling in air with supersonic speed and zero angle of attack are symmetrical about the axis of the missile. Surface bow and stern waves in water are symmetrical about a plane perpendicular to the water surface that passes through the axis of the model. As a starting point it may be assumed, therefore, that the model of the projectile should be constructed to be symmetrical about this plane and should be a projection of the original missile at a suitable scale.

Let us consider Eq. (3) as applying both to an artillery projectile traveling through air and, with subscripts, to a model of the projectile floating on a water surface. If we take the ratio of the projectile equation to the model equation, we have

$$\frac{D}{D_1} = \frac{C_D \rho v^2 A}{C_{D_1} \rho_1 v_1^2 A_1}. \tag{5}$$

Now if

$$f = f_1, \tag{6}$$

$$c/v = c_1/v_1, \tag{7}$$

$$\mu/\rho v L = \mu_1/\rho_1 v_1 L_1, \tag{8}$$

$$\alpha = \alpha_1, \tag{9}$$

and

$$L/\sqrt{A} = L_1/\sqrt{A_1}, \tag{10}$$

then

$$D = D_1 \rho v^2 A/\rho_1 v_1^2 A_1. \tag{11}$$

Equation (11) makes it possible to calculate the drag on a missile moving in air by making measurements on a model moving on the surface of a liquid, provided Eqs. (6) through (10) are satisfied to a reasonable degree of approximation.

## Experimental Investigation of Projectile Drag

At the present time the artillery projectile is the object most commonly encountered that travels with supersonic speed. The drag on such a missile is known by means of retardation measurements,[3] and the data may be used in checking Eq. (11). To make such a check a rather crude developmental supersonic wind-tunnel simulator was constructed (Fig. 1). In this device, water is pumped from tank $A$ to tank $B$ by means of two propellers in tube $C$, driven by a motor $D$. The excess water in tank $B$ flows down the open channel $E$ to tank $A$, thus sustaining the flow. The drag on a model $F$ is indicated by the deflection of the drag spring $G$.

[3] E. E. Herman, *Exterior ballistics* (U. S. Naval Institute, 1935).

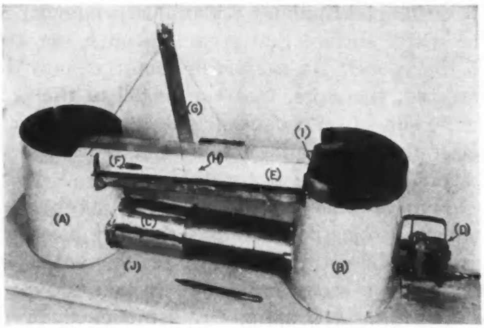



FIG. 1. Developmental hydraulic analog supersonic wind tunnel.

The deflection is measured from the reference line $H$ by a scale that is not shown. A series of parallel vanes $I$ are used to smooth out the surface of the water. The quantity of water flowing is adjusted by changing the speed of the pumping motor. The velocity of the water stream in the open channel is changed by adjusting the slope of the channel; this is accomplished by tilting the base $J$ of the simulator.

A model of a major-caliber artillery projectile that was used in making measurements is shown in Fig. 2($A$). The length is scaled down by a factor of about 1/70. The model is made of wood with a smooth lacquer finish, and its bottom is weighted with a thin sheet of lead cemented to the surface. The thickness of the lead was adjusted to give the immersion necessary to satisfy Eq. (10). When in the simulator, enough of the model projected above the surface of the water to prevent the bow waves from spilling over the top.

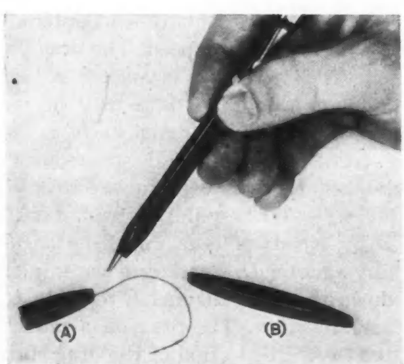



FIG. 2. ($A$) Model of a major-caliber projectile. ($B$) Model of a supersonic wing.

The first tests were made with zero angle of yaw to correspond to projectile retardation measurements taken under this condition. For this purpose a thread was cemented to the nose of the model about half way between the bottom and the water line. The thread was looped over a hook that projected into the water from the bottom of the drag spring. Thus the model drag was measued by the drag spring, and the depth of immersion was maintained by the model's displacement. The necessity for precise measurement of the relative position of the model and the water surface was therefore avoided. The drag was corrected for the component of the weight of the model in the direction of the water flow. A correction was also made for the drag on the tow thread. This was accomplished by using a zero drag reading established by measurements taken with a thread similar to the tow thread. Care was taken to select springs that operated within the linear range of deflection.

According to Eq. (7) the Mach numbers of the actual projectile and the model should be equal. The Mach number of the projectile is known because the velocity of the projectile and the velocity of sound when the retardation measurements were made are known. The Mach number of the model was determined by measuring the Mach angle[4] of the first bow wave in the supersonic wind-tunnel simulator. The angle that the bow wave makes with the axis of the model is

$$\theta = \arc\sin(c_1/v_1) = \arc\sin(1/M_1), \qquad (12)$$

where $M_1$ is the Mach number of the model. Figure 3($C$) is a picture of the model where the Mach angle was 43°, corresponding to a Mach number of about 1.5. This method of determining the Mach number of the model makes it unnecessary to know the velocity of surface-wave propagation. This is convenient because the velocity of propagation depends upon the wavelength, density, coefficient of surface tension and depth of the liquid. For open water that is deep

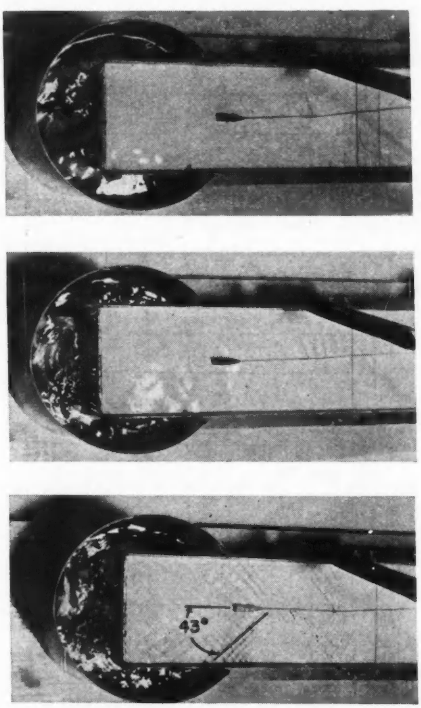
[4] Dodge and Thompson, *Fluid mechanics* (McGraw-Hill, 1937).

compared to the wavelength, the velocity of surface-wave propagation is

$$c_1=\left(\frac{2\pi T_1}{\lambda_1\rho_1}+\frac{g\lambda_1}{2\pi}\right)^{\frac{1}{2}}, \tag{13}$$

where $T_1$ is the coefficient of surface tension, $\rho_1$ is the density, $g$ is the acceleration due to gravity, and $\lambda_1$ is the wavelength. In Fig. 4, $C$ is plotted as a function of $\lambda_1$ from Eq. (13), for wavelengths less than 5 cm. In this region the velocity of water-wave propagation is approximately 0.001 that of sound in air. This reduction in the scale of speeds is very convenient.

It is interesting to note that the velocity of surface-wave propagation increases for wavelengths less than about 1.6 cm. This explains the shorter wavelength "capillary" waves that fan out in front of the main bow wave in Fig. 3($C$). These were disregarded in determining the Mach




(A)

(C)

FIG. 3. ($A$) Simulated subsonic flow. Water velocity less than the velocity of wave propagation. ($B$) Simulated transonic flow. Water velocity approximately equal to the velocity of wave propagation. ($C$) Simulated supersonic flow. Water velocity greater than the velocity of wave propagation. Mach angle = 43°, Mach number = 1.47.

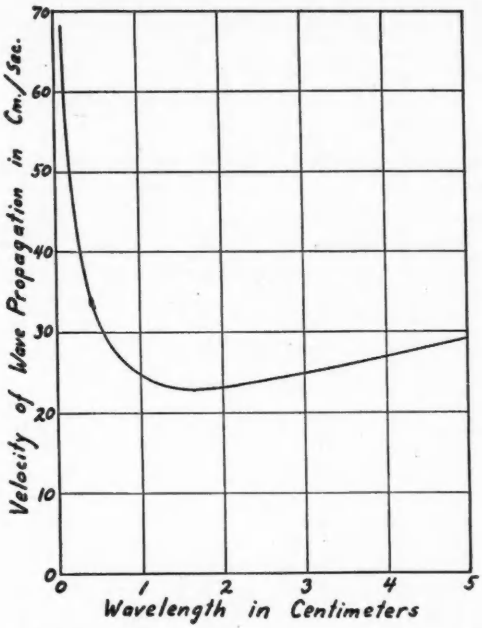



FIG. 4. Theoretical curve relating the velocity of water waves with their wavelength in deep open water.

angle. Figure 4 indicates that it would be desirable to use a simulator capable of handling models that produce bow waves of 1- to 4-cm wavelength because the velocity of propagation is relatively constant in this region. The simulator shown in Fig. 1 is not large enough for such models. The widths of the models shown in Fig. 2 are about 6 percent of the channel width. Wider models were tested, but there was a tendency for them to "choke up" the channel. This is discussed later.

For open water that is shallow compared to the wavelength, the velocity of surface-wave propagation is

$$C_1=(gh)^{\frac{1}{2}}, \tag{14}$$

where $h$ is the depth of the water. This interesting special case gives a velocity of propagation that is constant if the depth is held constant. The conditions encountered with the models used in the simulator shown in Fig. 1 gave waves whose velocity of propagation could not be expressed by the simple Eq. (13) or (14).

Equation (8) indicates that the Reynolds numbers should be equal for the missile and the model. This condition cannot be satisfied when water is used as the liquid medium. The Reynolds number for the projectile at $M=2$ was $R=0.85\times10^8$. For the model in water, $R_1=0.12\times10^5$ when $v_1=52$ cm/sec, corresponding to $M_1=2$. This discrepancy in the Reynolds numbers probably does not lead to serious error at velocities where waves are formed, because the wave resistance is a considerably more important factor than the viscous drag. It is interesting to note that the

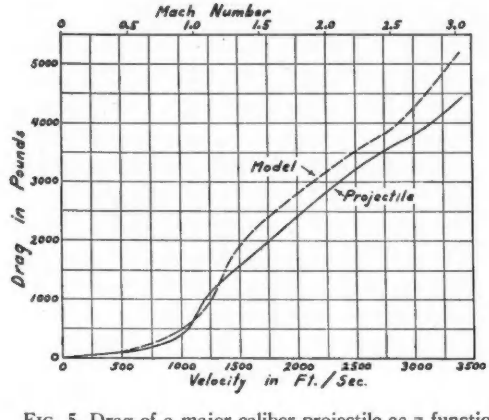



FIG. 5. Drag of a major caliber projectile as a function of velocity compared with the scaled up drag on a model measured in the simulator.

model in mercury would give $R_1=0.123\times10^7$, which is quite close to matching the missile in air.

In order to determine whether Eq. (6) holds true and at the same time to see how closely the other approximations are valid, the drag of the model was measured and compared with the known drag on the projectile, using Eq. (11). Figure 3 shows pictures of the model at water speeds corresponding to subsonic, sonic and supersonic speeds for the projectile. Figure 5 is a graph of the model drag, scaled up by Eq. (11), along with the projectile drag taken from proving-ground data for various velocities. It is evident that the two curves follow each other in general form and that the curve for the model is higher than the curve for the projectile, except in a narrow region near $M=1$. In the region $M<1$, this is probably partly due to the fact that the Reynolds number for the model is low compared to that of the projectile. For $M>1$, the discrepancy is probably due, in large part, to the fact that the Mach angle as measured for the model is smaller than the effective Mach angle. Figure 3($C$) shows that the Mach angle was measured to the first bow wave. Owing to the lesser capillary waves in front of the main bow wave, the effective Mach angle is probably greater.

Drag was measured for a model about twice the size of the one shown in Fig. 2($A$). When scaled up by Eq. (11) the measurements checked the model curve of Fig. 5 quite well in the region where $M<1.5$. Below this point the scaled up drag for the larger model was larger. It was found that "choking up" of the channel was at least partly responsible for this larger drag. When "choking" occurred, the water level immediately in front of the model would rise, causing the model to assume an angle with the horizontal considerably greater than the angle of the channel. Experiments on a model about three times the size of the one shown in Fig. 2($A$) intensified this effect.

Of the various factors in Eq. (11) that were used in getting the drag curve in Fig. 5 for the model, the projected area $A$ is probably known with least precision. This area was assumed to be the width of the model multiplied by the depth of immersion. The immersion was measured in still water and was approximately 5 mm; it could be read to about $\pm0.25$ mm. The velocity $v_1$ in Eq. (11) was measured by means of a Pitot tube placed at the position of the model after it had been removed. At low velocities the Pitot readings were checked by timing small pieces of cork as they moved a given distance in the region near the model. The spread of the experimental data is evident from Fig. 6. In this figure curve ($A$) shows the Mach number for the model plotted as a function of the velocity $v_1$ of the water relative to the model. Curve ($B$) is the model drag $D_1$ plotted as a function of the water velocity $v_1$.

It is of interest that at velocities corresponding to Mach numbers above 2.5 the position at which the tow thread was attached to the model was found to affect the drag and Mach number

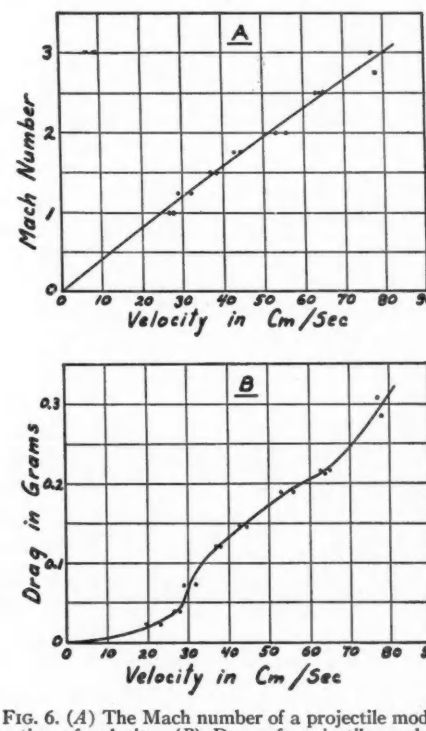



FIG. 6. ($A$) The Mach number of a projectile model as a function of velocity. ($B$) Drag of projectile model as a function of velocity.

readings. In this region the drag force becomes appreciable compared to the buoyant force, and the effective immersion is dependent upon whether the tow thread is attached high or low. For most accurate measurements above a Mach number of about 2.5, it is probably not desirable to depend upon buoyancy to maintain the depth of immersion constant.

Figure 7 shows the drag coefficients of the model and the projectile plotted as a function of the Mach number. The drag coefficients are defined by Eq. (3).

In order to be sure that the critical Reynolds number was not a major contributor to the shape of the model drag coefficient curve shown in Fig. 7, measurements of drag were taken on a submerged, thin model, the length of which was slightly greater than the model whose drag coefficient is shown in Fig. 7. The drag coefficient of the thin submerged model showed no rise for velocities corresponding to $M_1 > 1$. Thus the peak in the drag coefficient curve in Fig. 7 can be safely assumed to be caused by the formation of waves at $M_1 > 1$.

As indicated previously, the drag on an object such as a projectile increases as the axis of the projectile departs from the direction of the relative motion by an angle of yaw $\alpha$. The yaw also gives rise to a cross-wind force (called "lift" in aerodynamics) that acts at right angles to the drag force. The same dimensional analysis holds for the cross-wind force as was used in deriving the drag [Eqs. (2) through (11)]. It is only necessary to substitute $F$, denoting cross-wind force (or lift) for $D$, and $C_F$ for $C_D$. Thus Eq. (11) may be written:

$$F = F_1 \rho v^2 A / \rho_1 v_1^2 A, \tag{15}$$

where, it is remembered, Eqs. (6) through (10) must be satisfied to a reasonable approximation. There are very few reliable experimental data in the literature on the cross-wind force for supersonic projectiles at various yaw angles. For this reason no experimental comparative data for drag and cross-wind force for projectiles at various yaw angles are included in this paper.

## Experimental Investigation of Supersonic Airfoil Drag and Lift

Equations (3) and (4) can also be used to specify the drag on supersonic airfoils. For an airfoil, the characteristic area is taken as the projected area of the wing on the wing chord surface. If $a$ is used to denote this area, Eq. (3) may be written:

$$C_D = 2D/\rho v^2 a. \tag{16}$$

In like manner the coefficient of lift is

$$C_F = 2F/\rho v^2 a. \tag{17}$$

The drag and lift coefficients for several supersonic airfoils and for various angles of attack have been given by Ferri.[5] Airfoil GU-2 was selected for comparison with measurements made on a model in the supersonic wind-tunnel simulator. Figure 2($B$) shows the model of this airfoil. Figure 8 shows the model mounted in the simulator for angles of attack of 0°, 5° and 10°, everything else being held constant. The drag and lift spring is inserted loosely into a hole near the front of the model. This permits the model to maintain constant immersion by means of its buoyancy. A wire secured to the drag and lift spring is used to press against a vertical wire attached to the rear of the model. This creates a moment which maintains the angle of attack. The drag and lift were determined by measuring the components of deflection in the direction of, and at right angles to, the water flow. A scale (not shown in Fig. 8) was used to measure these deflections with respect to the transparent bridge across the channel, and to the side of the channel.

The coefficients of lift and drag for the GU-2 airfoil as taken from reference 5, and the coefficients as determined by measurements on the model in the supersonic wind-tunnel simulator, are shown in Fig. 9. It is interesting to observe

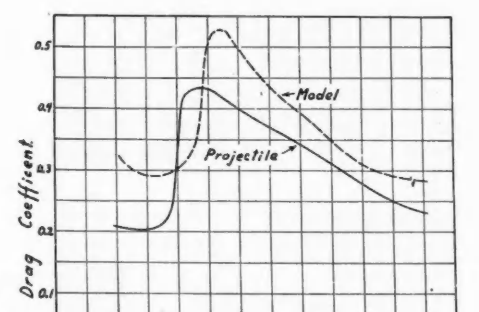
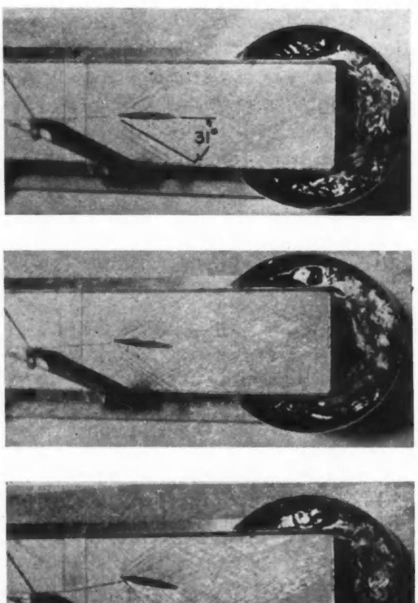



FIG. 7. Drag coefficient of a major caliber projectile as a function of Mach number compared with the drag coefficient of a model measured in the simulator.

[5] A. Ferri, "Experimental results with airfoils tested in the high-speed tunnel at Guidonia," Trans. in *NACA Technical Memorandum* 946 (1940).

that the drag coefficient for the model is high, as it was in the case of the projectile. It is also apparent that the lift coefficient for the model is considerably less than that for the GU-2 airfoil. In both cases the model curves are similar in shape to the airfoil curves.

It is not surprising that the numerical value of the lift coefficient is different for the model in the simulator than for an actual wing. The shock and bow waves for the wing and the model are farther from being comparable than for a projectile. The shock wave for a relatively long wing is formed at the leading edge and is the same throughout its length. For the model the "shock wave" is only on the surface of the water. Therefore, the immersion depth is not even approximately defined, as it is in Eq. (10) for a projectile. The depth of immersion of the model used was arbitrarily set at about 7.5 percent of its length. Experiments have shown that the drag and lift coefficients of the models are not critically dependent upon small changes of the depth of immersion. This does not mean, however, that the drag and lift coefficients are independent of the depth of immersion. It was not possible to explore this point fully




(A)

(B)

(C)

FIG. 8. (A) Simulated supersonic flow past an airfoil with 0° angle of attack. Mach angle = 31°, Mach number = 1.94. (B) Same with 5° angle of attack. Mach number = 1.94. (C) Same with 10° angle of attack. Mach number = 1.94.

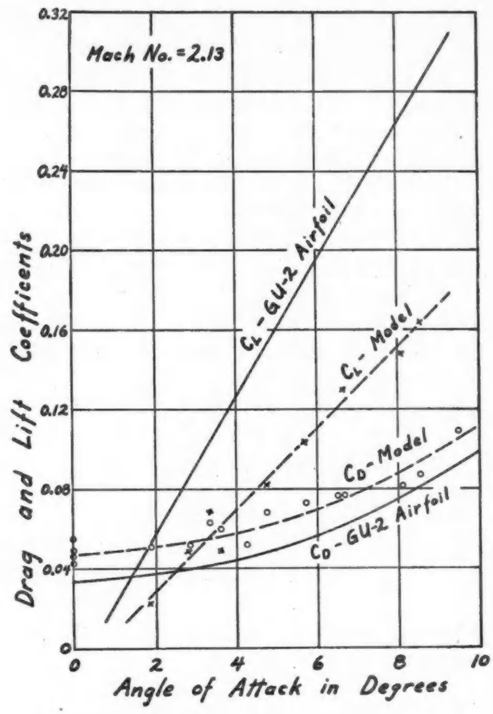



FIG. 9. Drag and lift coefficients of an airfoil at Mach number 2.13 for various angles of attack compared with the coefficients of a model measured in the simulator.

with the developmental simulator used because of the limited water depth (8 mm) in the channel due to insufficient pumping capacity. If a greater immersion depth were used, the coefficient of lift curve for the model in Fig. 9 might be higher and the coefficient of drag curve might be lower. There is a possibility that by adjusting the immersion, the curves for the airfoil and the model could be matched more closely. This interesting investigation will be pursued when a larger simulator is available.

Moment measurements were taken by measuring the deflection of the wire creating the moment couple. These measurements were not accurate, owing to the lack of mechanical refinement, and are not given here. The measurements did indicate, however, that the moment coefficient of the model was low compared to that of the GU-2 airfoil. This is expected because the lift coefficient is low.

## Comments and Conclusions

The afore-described experiments indicate that measurements of lift and drag on models moving relative to a liquid surface can be used in determining the approximate drag and lift on projectiles and wings traveling through air at

supersonic speed. The drag coefficients for the models tested were found, for the most part, to be higher than the coefficients of the actual projectile or wing. The lift coefficient curve for a model of a supersonic wing was found to be lower than the curve for the actual wing. In all cases the coefficient curves for the models were similar in shape to the coefficient curves of the actual projectile or wing. Thus it appears that more accurate measurements can be made by applying correction factors determined by spot checks using actual supersonic wind-tunnel data.

By the use of the supersonic wind-tunnel simulator, approximate drag and lift measurements on projectiles and wings can be made on a test bench with the aid of inexpensive equipment. This opens the way for more students to gain experience in supersonic flow. It is emphasized that the measurements are considered to be more qualitative than quantitative.

The supersonic wind-tunnel simulator used for the experiments was a handmade developmental model. The simulation and the mechanical accuracy of the apparatus are not perfect, as will be evident to the precisionist. For more precise measurements the simulator should be mechanically refined and made larger. This would permit the construction of larger, more accurate models and make possible more accurate measurements. Capillary waves of relatively high velocity should be less noticeable if larger models are used.

Other measurements, in addition to those for drag and lift, can probably be made, although they have not been thoroughly investigated. Rough preliminary checks indicate that approximate moment measurements can be made with the simulator. The approximate pressure distribution around a model could probably be observed by using Pitot tubes built into the model. By the same method the approximate pressure distribution inside a model with an internal duct could probably be studied. Air or water jets built into a model can be used to simulate jet propulsion.

# A Square-Wave Generator for Instructional Use

E. H. Green and W. H. Mais
*Brooklyn College, Brooklyn 10, New York*

WE have constructed a square-wave generator for use in our advanced laboratory which differs in some respects from those commercially available or previously described.[1] The unit has been designed so as to facilitate the study of the method of production of square waves and is arranged for convenience in studying the distortions of square waves that are produced by typical networks.

The apparatus consists of a multivibrator section, a clipping section, an amplifier, and an output cathode follower, as shown in Fig. 1. The entire apparatus, including a power supply unit, was built on a chassis about 8×12×3 in.

The amplifier triode serves two purposes. First, it increases the amplitude of the voltages fed to it at *E*. Thus the oscillograph amplifiers, which usually have an inadequate range of response for such purposes, need not be employed. Second, the amplifier's grid is never driven to cut-off voltage. It therefore presents a resistance to the test network that does not vary throughout the cycle, and which may therefore be included in the network computations.

The test networks are rigidly mounted on Bakelite slabs, about 3×4 in., and plugged in to make connection at *F*, *G* and *H*. The banana

[1] H. N. Walker and P. Greenstein, "Direct current transients with the square-wave generator," *Am. J. Physics* 10, 198 (1942).

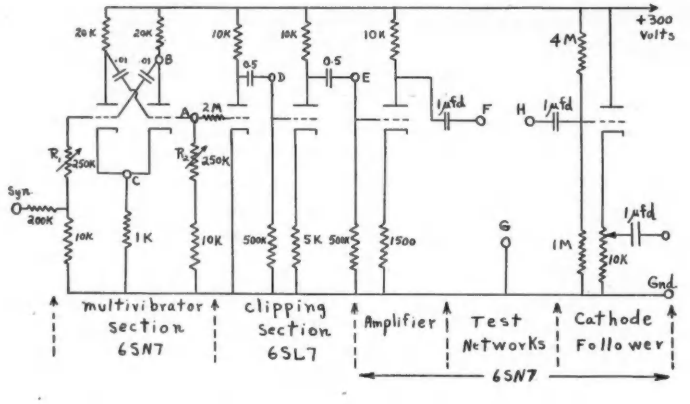



FIG. 1. Wiring diagram of apparatus.

outlets $F$, $G$ and $H$ are spaced so that the network mount can be inserted in only one orientation. The output of the test network is fed to a cathode follower, which has the property that its output resistance is only a few hundred ohms. The rigid mounts and the cathode follower prevent distortions of the patterns due to the presence of shunt capacitances of the students' lead wires or of the internal oscillograph connections.

Since the precision of the comparison of $RC$ values with the decay time observed on the oscillograph screen is necessarily limited, the values of the network elements need be known only approximately, and the resistors may be of the carbon type.

The frequency of the square waves may be varied from about 150 to 2000 c/sec by means of the ganged resistors $R_1$ and $R_2$. A calibrated dial gives the approximate values of the frequency directly. Various test points are provided at $A$, $B$, $C$, $D$, $E$ or $F$, by bringing leads from these points to banana outlets on the top of the chassis. The output cathode follower may then have its input terminal $H$ connected to any of these test points,

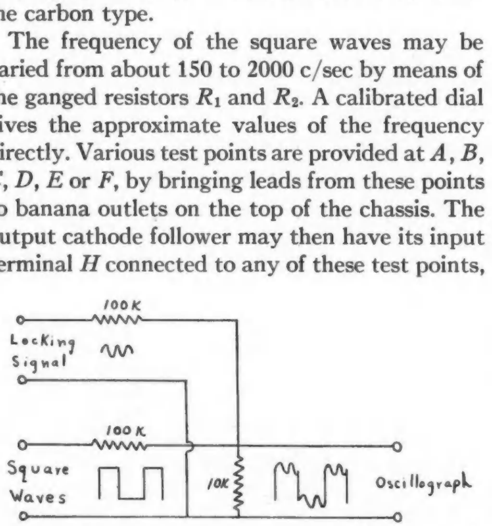



FIG. 2. Method of studying "locking-in" process.

and the output voltage from the cathode follower can be studied with a cathode-ray oscillograph.

The student observes the multivibrator voltage wave form at the grid $A$ and at the plate $B$. An observation at the common cathode connection $C$ shows the double frequency of this voltage. Connecting to the points $D$ and $E$ shows the clipping actions produced in these two stages. A synchronizing terminal is provided so that a synchronizing voltage can be injected into the grid circuit of the multivibrator section. The "locking-in" process may then be studied by adding a fraction of the synchronizing voltage to the square-wave output (Fig. 2). The resulting pattern is easily interpreted.

The student is asked to account for the wave forms observed with each of the following circuits, chosen to illustrate some of the more common or particularly instructive networks.[2]

(1) *Differentiating networks.*—The network of Fig. 3($a$) is arranged with $RC \approx 10^{-3}$ sec. The multivibrator frequency is varied so that the period of the square waves ranges from a value much larger than $RC$ to one much smaller than $RC$. The differentiating action is easily traced with these circuit constants. The circuit also shows the effects of poor low-frequency response in a resistance-coupled amplifier. To demonstrate very narrow pulses and a typical "pulse-sharpener," we have found the networks of Fig. 3, ($b$) and ($c$), to be suitable.

(2) *Integrating networks.*—These networks [Fig. 4, ($a$) and ($b$)] incidentally show the effects

[2] Some sample patterns are reproduced in reference 1.

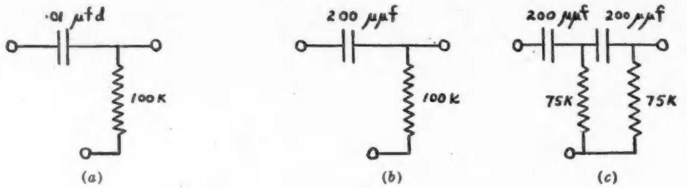



FIG. 3. Differentiating networks.

of poor high-frequency response in a resistance-coupled amplifier.

(3) *Compensated voltage divider*.—This circuit (Fig. 5) is encountered in practice in the following form. A voltage divider $R_1$, $R_2$ at the input of an amplifier whose input capacitance is $C_2$ is equipped with an additional capacitor $C_1$ adjusted so that $R_1C_1=R_2C_2$. (Examples are found in the Dumont Type 241 oscillograph and the Ballantine vacuum-tube voltmeter.) The students' adjustment of $R_1/R_2$ changes the pattern from "differentiation" to "integration." The correct setting for compensation is easily determined.

(4) *Compensation for inductive reactance*.—If $R_1=R_2=\sqrt{(L/C)}$, the impedance of the network (Fig. 6) is independent of frequency and equal to $\sqrt{(L/C)}$. An example of the application of this principle is found in compensation for the inductive reactance of the voice coil of a loudspeaker. Convenient values of the circuit constants are: $L=1h$; $C=0.01$ μf. The resistors can be about 25,000 ohms, variable.

(5) *Transient oscillations*.—Use of the oscillograph amplifiers will not be objectionable in this case (Fig. 7). Two sets of transient oscillations will be observed, due, respectively, to the positive and negative portions of the square wave. These may be separated by vertical displacement, if desired, by insertion of a variable resistance at $X$. The measured values of $L$, $R$ and $C$ provide data for computation of the expected decrement of the oscillations, as well as of the frequency. Depending upon the desired precision of the comparison of computation and observation, the frequency determination may be made either by counting the number of oscillations during the known half-period of the square waves, or by forming a 1:1 Lissajous figure with the output of a laboratory oscillator. Despite the decay of the oscillations, the Lissajous figure will be readily recognizable, especially if vertical separation is employed. We have found that the precision of the comparison is such as to justify the accurate determination of $L$, $C$ and $R$ by means of a bridge. The damping effect of the 1-megohm resistor, which is effectively in parallel with the transient network, must be included in the computation, for full precision.[3] If it is not desired to bring out this point, the resistor should be replaced by a capacitor of about 200 μμf. This will contribute a negligible damping.

(6) *Characteristics of an interstage transformer*.—An interstage audio transformer, used as a test network (Fig. 8), shows characteristically a slope of the horizontal portion of the square wave, owing to failure to respond to very low frequencies, and a transient oscillation due mostly to the distributed capacitance and leakage inductance of the coils. These transformers are ordinarily designed to operate into a resistive secondary load of several hundred thousands of ohms. We, therefore, provide a snap-button con-

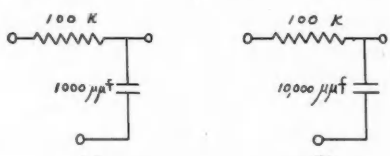



FIG. 4. Integrating networks.

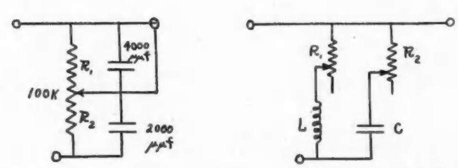



FIG. 5 and 6. Left, compensated voltage divider; right, compensation for inductive reactance.

[3] Morecroft, *Principles of radio communication* (Wiley, ed. 2), chap. III, Eq. (13).

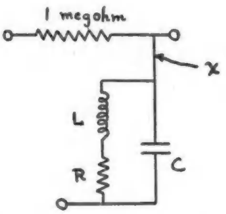



FIG. 7. Transient oscillations.

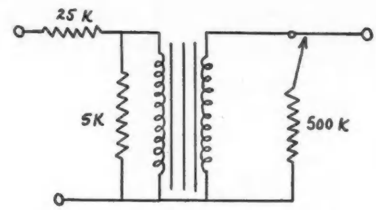



FIG. 8. Interstage audio transformer.

nector that places a 0.5-megohm resistor across the secondary when desired. The transient oscillation is thereupon much reduced in amplitude. The behavior thus qualitatively observed is compared with that to be expected from a point-by-point graph of the frequency response of this transformer, which is found to show a sharp rise at the frequency of the observed transient. The voltage divider shown at the primary side of the transformer serves to reduce the applied voltage, to prevent overloading of the cathode follower.

The study of transient oscillations in inductively coupled circuits tuned to the same resonant frequency is carried out in this laboratory by the procedure described earlier by one of us.[4] However, a suitable coupled-circuit network could readily be constructed for use with this equipment. Recently released material on radar technics may suggest other useful networks to the reader.

[4] E. H. Green, "A simple arrangement for observation of electrical transients," *Am. J. Physics* (*Am. Physics T.*) **5**, 181 (1937).

## Potential Difference in Textbooks for Beginners

FRANCIS W. COOKE
*State Teachers College, St. Cloud, Minnesota*

ELECTRIC potential difference is a concept that is considered to be difficult for students to understand and that often is not clearly defined in textbooks for beginners.

In June, 1938, a report[1] of the Association's Committee on Electric and Magnetic Units suggested the following definition:

> The difference of potential between two points $A$ and $B$ is the quotient obtained when the work done in moving the electricity $q$ from $B$ to $A$ is divided by $q$; $V=W/q$. Or, it is the quotient obtained when the power required to maintain a current between $B$ and $A$ is divided by this current; $V=P/i$.

A recent examination of 42 textbooks on general college physics, published in the United States from 1935 to 1945, indicates that potential difference is seldom as carefully defined as in the definition just cited. The results of the study of the 42 textbooks are summarized in Table I. Only the latest edition of each book is listed. As the table shows, the aforementioned report seems to have had little effect on the definition of potential difference in the books published after 1939, as compared with those published in 1939 and before. The table shows also that the books that have been revised one or more times do not give any more satisfactory treatment than the others. This is seen to hold also for those textbooks of which the latest revision appeared after 1939.

Most textbooks, of course, state that potential

[1] *Am. J. Physics* **6**, 144 (1938).

difference is related to work, but the majority do not state clearly the exact relation. There is naturally much variation in the phrasing of the definitions in the various textbooks, but it was found that the definitions could be classified in four groups as follows.

(i) *Incorrect statement.* Potential difference is defined as the work done on a unit charge, or as the potential energy possessed by a unit charge, with no mention of the ratio of the work to the charge on which the work is done.

(ii) *Indefinite statement.* Potential difference is discussed, with or without the use of analogy, but no definite statement is given as to what potential difference is; or only a unit of potential difference is defined.

(iii) *Contradiction.* Potential difference is defined as work; and also as work per unit charge, or by the equation $V=W/Q$.

(iv) *Satisfactory definition.* Potential difference is defined as the ratio of work done on a charge to the charge; or as work done per unit charge; or by the equation $V=W/Q$, with the symbols defined; or in an equivalent statement; and without contradictory or indefinite statements.

When a student reads in his textbook that the potential difference between two points is the work done on a unit quantity of electricity in transporting it from one of the points to the other, he may naturally wonder why this work, expended in this particular way, is given a new name, *potential difference.* In the student's previous experience, work has always been merely work, no matter how it was used.

Many of the quotient concepts of physics are defined as quotients. For example, the density of a substance is usually defined, not as the mass of a unit volume of the substance, nor as the mass or volume of anything, but as the quotient obtained when the mass of a piece of the substance is divided by the volume of that piece. At least 33 of the 42 textbooks examined define density in this way. Likewise, at least 30 of the textbooks examined define capacitance as the quotient of quantity of electricity and potential difference, and without contradictory statements, and 23 of them actually use the word *ratio* in giving the definition. In contrast, only 11 of the

TABLE I. Definitions of potential difference.

| Year of publication | Number of books | Number of books in which definition is: Incorrect | Indefinite | Contradictory | Satisfactory |
|---|---|---|---|---|---|
| | | *New Books* | | | |
| 1935–39 | 13 | 3 | 2 | 4 | 4 |
| 1940–45 | 13 | 3 | 3 | 2 | 5 |
| | | *Revised books* | | | |
| 1935–39 | 8 | 1 | 1 | 5 | 1 |
| 1940–45 | 8 | 3 | 0 | 4 | 1 |
| | | *All books* | | | |
| Total | 42 | 10 | 6 | 15 | 11 |

textbooks are classed as giving satisfactory definitions of potential difference, and only one of these uses the word *ratio* in the definition. It is unfortunate that potential difference is so seldom defined as a quotient in the textbooks for beginners.

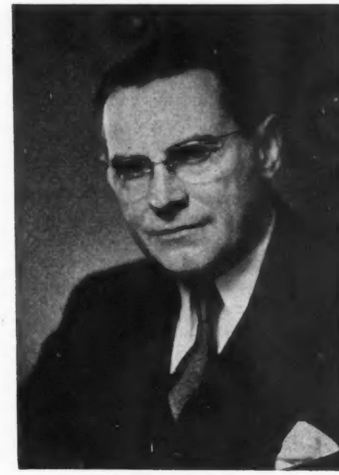
The case for defining a quotient as a quotient is well stated by Howe[2]:

> The term "per unit" appears in so many definitions of physical quantities that its true significance should be kept clearly in mind. Any quantity defined by the use of this phrase is a *quotient.* For example, "mass per unit volume" means the quotient of any particular mass by the volume that mass occupies. This quotient is neither a mass nor a volume, and it cannot be expressed in either mass units or volume units. It is a new quantity, named *density*, and must be expressed in a derived "per" unit, such as grams per cubic centimeter. . . . Similarly, speed is *not* "the distance gone in unit time"; speed is "distance per unit time," a quotient.

It is therefore suggested that textbook writers, in defining potential difference, should be guided by the report of the Committee on Electric and Magnetic Units; that they define potential difference as a quotient, obtained only by division; and that they use the words *quotient obtained* and *divided by*, or equivalent words, prominently in the definition.

[2] H. Howe, *Introduction to physics* (McGraw-Hill, 1942), p. 138.

*You don't have to preach honesty to men with creative purpose. Let a human being throw the engines of his soul into the making of something, and the instinct of workmanship will take care of his honesty.*—WALTER LIPPMANN.



**Duane Roller**

## Recipient of the 1946 Oersted Medal for Notable Contributions to the Teaching of Physics

*The American Association of Physics Teachers has made to Professor Duane Roller, of Wabash College, the eleventh of its annual awards for notable contributions to the teaching of physics. The addresses of recommendation and presentation were made by Professor Lloyd W. Taylor, Chairman of the Committee on Awards, and Professor R. C. Gibbs, President of the Association, in a ceremony held in McMillan Theatre, Columbia University, on January 31, 1946, during the sixteenth annual meeting.*

### INTRODUCTORY REMARKS BY PROFESSOR R. CLIFTON GIBBS, PRESIDENT OF THE ASSOCIATION

THE Oersted Medal, established for the purpose of giving public recognition to persons who have made "notable contributions to the teaching of physics," was first awarded by the American Association of Physics Teachers in 1936.

At each succeeding annual meeting of the Association this means of signalizing the achievements of a worthy devotee of this most honorable calling has occupied a prominent place in our programs. We have sought to discover and to place a mark of distinction, an "order of merit," upon a few of our colleagues who have contributed outstandingly to the practice of a difficult art, the teaching of physics.

These and other allied purposes were, I am sure, the actuating mainsprings in the minds of those in our organization who were responsible for initiating the concept of an award of this kind and for seeing it established. To one man especially much credit is due for those early efforts. To PAUL E. KLOPSTEG we extend our warm appreciation for the constructive support he gave to this project from the beginning, not only in the way of time and thought but also in seeing that adequate funds were provided.

This year we make our eleventh annual award of the Oersted Medal to a man who, because he is a former student of previous medalists, can be considered as the latest link in a chain reaction to which we may confidently expect other links will later be added. Certainly the opportunity and the need for making a notable contribution to the teaching of physics was never greater than now, as we enter upon a period of unprecedented but as yet largely uncharted expansion of our subject.

With our thoughts thus looking forward hopefully into the future, we pause to examine and to take stock of one of our present outposts and so call upon PROFESSOR LLOYD W. TAYLOR, Chairman of the Committee on Awards, to present the Medalist for 1946.

## ADDRESS OF RECOMMENDATION BY PROFESSOR LLOYD W. TAYLOR, CHAIRMAN OF THE COMMITTEE ON AWARDS

HERETOFORE the identity of the Oersted Medalist has—in theory, at least—been kept secret until the moment of the actual award. Last year the Executive Committee directed that henceforth public announcement should be made well in advance of the meeting at which the award was made. This has been done;[1] this year's recipient is DOCTOR DUANE ROLLER, of Wabash College.

The Oersted Medal is awarded annually by this Association "for notable contributions to the teaching of physics." Needless to say, such contributions may be made in a great variety of ways and the awards of the last three years illustrate this variety. Last year's award was to a man who, working virtually alone and in relative professional obscurity, had yet produced a record number of outstanding physicists. The year before the award was to a man who, necessarily working more in the limelight, had done more than any other one man to build up the *esprit de corps* of college teachers of physics and to vitalize the channels through which they could cooperate professionally to maximum effect. This year the chief basis of the award, although unlike either of the two preceding, is yet supplementary to both. It is the establishment and maintenance of an effective channel of rich communication between teachers of physics.

In the course of our correspondence on this award DOCTOR ROLLER wrote, "I suppose that what little reputation I have has been gotten chiefly in my capacity as an editor." Except for the first qualifying phrase, how right he was! He has been the editor of the *American Journal of Physisc* from its first issue in 1933 under the name of *The American Physics Teacher*. The 77 subsequent issues constitute a contribution to physics that any man could be proud to cite as his life work. Of all "notable contributions to the teaching of physics," the *Journal* is one of the most outstanding.

Perhaps the first and greatest contribution to the teaching of physics, subsequent to its first appearance in college and university curriculums, was the introduction of laboratory work into the instructional program. One of the early Oersted awards was to a man who had had a considerable hand in this phase of our development. I should like to state my own humble opinion that, next to the introduction of laboratory work, the greatest contributions to the teaching of physics at the college and university level have been made through the columns of the *American Journal of Physics* under the active (some of you would be tempted to say belligerent) editorial guidance of DUANE ROLLER. Aside from acting as a particularly potent watchdog in all matters of style, form and content of articles submitted for publication, he has himself contributed several hundred digests of periodical articles, book reviews and editorial notes. The intimate imprint of his personality is apparent in every one of the 4710 pages of the *Journal* to date, invariably to the advantage of authors and the improvement of the *Journal*.

Still more important than the maintenance of a high standard of journalistic style has been the fostering of a broad and comprehensive conception of the place of physics in our modern culture. As DOCTOR ROLLER has himself formulated the policy of the *Journal:*

> It provides material for improving all aspects of physics education, including the instruction of students who are specializing in physical science and of those who study physics as part of a liberal education. Its articles and notes are intended to assist and encourage teachers in all types of institutions to keep abreast of the state of the science, and to continue in creative work of one kind or another. It seeks to promote a better understanding of the objectives and the difficulties inherent in the various borderland fields of the science, and to enhance the realization that the philosophic, historical and socio-economic aspects of physics are integral parts of the science, the science itself being an integral part of our modern culture.[2]

DOCTOR ROLLER once remarked that the *Journal* was a monument to broken Sabbaths. The writer ruefully recalls, indeed, how one Sunday afternoon several years ago, MRS. ROLLER was compelled to remind her husband of the

[1] *Am. J. Physics* **14**, 447 (1946).

[2] AAPT bulletin, "History and activities . . . " (July 1946), p. 13.

urgency of some accumulated editorial work while out-of-town guests were present at the Roller home. That occasion was doubtless typical of the imperious demands that the *Journal* has unremittingly made on the time of DOCTOR ROLLER. Nor should it be overlooked that MRS. ROLLER herself has been an able editorial assistant and "critic on the hearth." The Committee hopes that the membership will regard her as sharing this award.

Also it will possibly not be inappropriate to record that DOCTOR ROLLER has never received a cent of compensation for his work on the *Journal*, although similar journals pay sums running into four figures for services no more valuable than his have been. In its earlier years the Association simply did not have the wherewithal to pay an editorial stipend. In more recent years DOCTOR ROLLER has declined to receive such a stipend. It has been possible just lately, however, to provide him with a certain amount of competent editorial assistance. Also, he has been persuaded to receive free passage, accompanied by MRS. ROLLER whenever possible, to meetings of scientific societies other than this Association. This is not so much by way of compensation as to enable him to broaden his contacts with the world of physics and lay the foundation for a still richer *Journal*, if such a thing is possible.

DOCTOR ROLLER's contributions to physics, and more especially the teaching thereof, have not been confined to the editorial channel. He is a co-author of three books, with a fourth just over the horizon, and has written some 40 major articles for the profession. In addition, he has built up a reputation as an unusually skillful classroom teacher and supervisor of teachers, and has pursued some professional hobbies in fields which some of us, if we do not take them entirely for granted, seldom do more than "gripe" about. One of them he will presently describe to us, under the title of "An Approach to the Study of Physical Terminology."

For ten years DOCTOR ROLLER was an ex-officio member of the Committee on Awards, all other members of which were steadily changing with changes of officers. A mounting conviction on the part of the other members that an award to him was long overdue posed the problem of how to push him off to clear the way for such an award. Last year's committee finally took the bit in its teeth and proposed its own reconstitution, leaving out the Editor. The pretext was that ten years was long enough for one man to be on any committee, that if he had not made his contribution in that length of time, it was unlikely that he ever would. The proposal met with favor, and DOCTOR ROLLER was unceremoniously dumped overboard.

*Mr. President, I am deeply gratified, on behalf of the Committee on Awards, to present* DOCTOR DUANE ROLLER *for receipt of the eleventh Oersted Medal, awarded annually by this Association "for notable contributions to the teaching of physics."*

## An Approach to the Study of Physical Terminology

DUANE ROLLER
*Wabash College, Crawfordsville, Indiana*

*The use of language in science is specialized and peculiar. In a brief speech the scientist manages to say things which in ordinary language would require a vast amount of talk. His hearers respond with great accuracy and uniformity. The range and exactitude of scientific prediction exceed any cleverness of everyday life: the scientist's use of language is strangely effective and powerful. Along with systematic observation, it is this peculiar use of language which distinguishes science from nonscientific behavior.*—LEONARD BLOOMFIELD.[1]

THE role of language in a science is of utmost importance; for not only is communication of ideas indispensable if we are to have any science, but the symbology and framework of language used in this communication are also the very tools with which we *think*. We do not think first, and then afterwards translate the results into words. Instead, clear thinking and the correct use of words are essentially the same process.

The experienced physical scientist is not likely to feel much concern about questions of faulty

[1] L. Bloomfield, *Linguistic aspects of science* (Univ. of Chicago Press, 1939), p. 1.

terminology so long as he is dealing with the older and better established parts of his field. He has become accustomed to such terms as "gram molecule," "vapor tension," "permanent gas," "latent heat," "chemical ray," "heat radiation" and "electromotive force." To him these are familiar friends, with histories that clearly account for their existence. The total number of earlier concepts is comparatively small; and each old term, no matter how faulty it may be, is for the trained person merely an established symbol, or label, for a definite idea. It lacks the flexibility of connotation that is characteristic of words in general. Thus, in the older parts of physical science, it is not the mature scientists but students and many other comparatively untrained users of technical terms who are most likely to benefit from reforms in terminology.

This is less true when the field in question is comparatively new, or is becoming highly complicated, or is developing rapidly in separate research centers located in different parts of the world. Perplexing problems of language are then likely to arise in the realms of research. Nor is it true of those parts of physics, old or new, that have important industrial and commercial applications; for obviously it is essential that intelligible and unambiguous language be used in framing commercial specifications, contracts, and descriptions of patents. Without much question, it is the physicists in industry, rather than those connected with colleges and universities, who are today generally most sympathetic toward proposals for improving technical language and who are most active in instigating and pursuing studies leading to such improvement.

Obviously the opportunities offered by a study of terminology transcend those of commercial significance, important as the latter may be; and certainly the results of such a study can be much more than the mere replacement of a few old terms by others that are less bothersome. This is especially true in the newer fields, where the necessity for naming brand-new concepts is continually arising and where the existing nomenclature is likely to be already somewhat logical in structure or, in any event, less firmly established. Not only is the chance better that reforms in nomenclature can be effected, but also that the conceptual structure of the field can be improved.

There is an additional, somewhat different incentive for improving technical language. As an analysis of physical literature clearly shows,[2] the best of our modern research papers and treatises are so obviously effective as to make certain that in no other fields of knowledge are ideas and meanings recorded and transmitted with so much exactness and clarity. The physical scientist has made tremendous contributions to language and communication; but, because he regards these contributions as secondary products of his work in physical science, he has not allowed himself to be sufficiently impressed by them. As was said earlier,[2] "we should recognize that many problems of language lie within our province and should feel freer than we have in the past to participate in language reforms when they appear to be needed."

In the study of physical terminology, the task is not merely to record usage but, when it seems desirable, to attempt to create usage. This means not that we can supplant experts in linguistics, but that their problems and ours overlap to a greater extent than is commonly supposed. It also means that improvements in physical terminology can have important effects, not only within the sciences, but on the art of communication as a whole.

Nevertheless, since the primary reason for making such improvements is to expedite scientific work, the question arises as to who is in the best position among physicists to carry out reforms in language. Clearly it is the workers in any particular, highly specialized field of research who can most effectively assess the language needs of that field, and who should assume primary responsibility for language reforms within their own areas. Yet no one is likely to contend that the nomenclature of any specialized branch of physics should differ in principle or be in conflict with that of physical science as a whole. Any tendency in separate research areas to develop completely independent nomenclatures would have a retarding effect on the integration and consequent simplification of knowledge that are essential if these separate fields are in the long run to progress most rapidly.

---

[2] See, for instance, D. Roller, "Technical writing and editing," *Am. J. Physics* **13**, 99 (1945); also, reference 1.

Nor are the requirements of the highly creative research worker the only ones to be kept in mind. With the demands for scientific personnel rapidly increasing, it is no longer possible to rely solely on first-class minds to carry out all the tasks of research; provision will have to be made for utilizing many persons of lower ability who can still do useful work under direction. With the average ability of scientific workers thus decreasing while knowledge is growing in complexity, the need is greater than ever before for the utmost clarity of language and thought.

There is still a third group to be considered. E. Q. Adams,[3] in speaking of chemical terms, says that the primary purpose of these terms is neither that of being used by research workers nor that of being taught to beginners in chemistry, since the vast majority of those who use them are neither one nor the other; nevertheless, he adds, it is on these two frontiers of knowledge—the one where it encounters new facts, the other where it meets new minds—that the serviceability of a term is most searchingly tested. There is then a pedagogic as well as a purely technical aspect to nomenclature. No terminology is satisfactory if it fails to expedite the training of each new generation of scientists. Nor will it then be highly satisfactory for the research worker himself. If, in the extreme case, all technical terms were chosen with complete disregard for ease in remembering their meanings and the relations among them, even a trained worker would have to depend continually on a technical dictionary.

### Suggested Steps in the Study

Some 24 centuries ago, Myron, a Greek sculptor, made the observation that one ought not to investigate things from words, but words from things. This would seem to be a sensible procedure. Yet even today it is mainly in the sciences that the fertility of this approach has been recognized. Applied to a systematic study of terminology it implies that the name of a concept should not be chosen until after the concept has been defined as carefully as is possible. This is one reason why a thoroughgoing study of terminology can be quite profound in its outcomes; for in reexamining and possibly redefining ideas, there is presented the opportunity to generalize, reclassify and integrate, and so obtain simpler and more logical hierarchies of concepts.

The concepts that have to be defined and named include:

(i) Active physical concepts;
(ii) Devices and apparatus;
(iii) Passive properties of devices, bodies and substances;
(iv) Units;
(v) Processes and results of processes.

In addition to the framing of a satisfactory definition, it is necessary to compile a list of all the synonyms for the concept that are in current use. Then comes the task of selecting the single best name in the light of criterions set up for this purpose. If none of the existing names proves to be satisfactory, or if the concept is entirely new and as yet unnamed, a suitable new term must be found.

### Criterions for Selecting Terms

It seems reasonable to assert that an ideal physical term is one that satisfies five basic requirements: nonambiguity, meaningfulness, internationality, simplicity and euphony.[4] An examination of many existing terms in the light of these general requirements indicates that the following somewhat more specific rules are useful in selecting terms and constructing a technical glossary.

As will be seen, these rules are practical ones, being intended for immediate use in effecting day-by-day improvements. They serve to systematize the procedure that we have been using in the past; namely, of refining and patching up ethnic languages so as to make them more useful for the purposes of physical science. Ordinary languages of course had their origin in prescientific thought and hence have characteristics that may seriously retard very fundamental investigations; perhaps no amount of patching up of such languages will render them suitable for certain kinds of thinking. However, only a few scientists are likely to be affected by such a breakdown of language for a long time to come. For the vast majority the practical rules should prove to be the most useful.

#### *1. A given term should have only one technical meaning*

If, for instance, *power* is the accepted term for time-rate of transferring or transforming energy,

[3] E. Q. Adams, *J. Chem. Ed.* **1**, 230 (1924).

[4] Except for "meaningfulness," these are the requirements listed by P. Moon, "The names of physical concepts," *Am. J. Physics* **10**, 134 (1942).

then one must regard as faulty such terms as absorptive power, candlepower, magnifying power, resolving power, rotatory power, thermoelectric power. Satisfactory equivalents for most of these terms already exist and should be adopted exclusively.

Similar considerations apply to a number of other terms—for example: absolute, action, center of gravity, d'Alembert principle, density, dichroism, dispersion, equilibrium, intensity, Kerr effect, Kossel-Sommerfeld law, mass, power factor, specific, tension, transmission coefficient, viscosity, Wiedemann effect, Wollaston prism—all of which have two or more distinctly different meanings in physical science.

### *2. A given concept should have only one name*

There are several hundred physical concepts that have two or more names in current use. Among those having at least four names are:

angular momentum, moment of m., spin m., rotational m.

atom form factor, atomic scattering f., atomic structure f., structure amplitude f., *f*-value.

autoelectronic emission, cold e., field e., field current.

bending torque, flexural t., bending moment, flexural m.

blackbody radiation, cavity radiation, black radiation, Planckian radiation.

capacitor (device), condenser, capacitance, capacity.

electric field strength, e. field intensity, e. field, e. intensity, e. force.

electric strength, dielectric s., disruptive s., critical gradient, puncture voltage.

electrochemical constant, electrolytic c., Faraday c., Faraday electrochemical c., Faraday electrolytic c., faraday.

kilocalory, Calory, kilogram calory, large c., great c.

magnetic quantum number, equatorial q.n., axial q.n., *M* value.

mole, gram mole, gram mol, gram-molecular weight, molar w., molal w., gram molecule.

noncircuital vector field, lamellar f., irrotational f., noncyclical f.

powder pattern, Debye-Scherrer circle, Debye-Scherrer ring, Hull r.

shear modulus, rigidity m., torsion m., slide m., Coulomb's m., m. of transverse elasticity, rigidity, coefficient of simple rigidity.

volume modulus, bulk m., m. of cubical compressibility, m. of compression, resistance to compression, coefficient of volume elasticity, cubical elasticity.

Young's modulus, stretch m., longitudinal m., m. of traction, m. of direct elasticity, coefficient of resistance to extension.

In the field of electrical engineering alone there are at least 200 concepts that have two or more names in current use.[5] If a critical study of terminology resulted in nothing other than the dropping of needless synonyms, we would benefit by having our total working vocabulary in physics proper reduced by at least 15 percent, or, roughly, 1000 terms. Thus the time saved by not having to learn such terms would be appreciable. Moreover, the resulting reduction in vocabulary would be a step forward in making it easier to adopt English as the secondary world language for physics.[6]

It is true that many of these superfluous terms—"latent heat" is a good example—have considerable historical interest and hence should be mentioned in historical accounts. But such terms should be clearly designated as obsolete so far as the current working vocabulary is concerned.

It is worthy of note that only in the sciences do *exact* synonyms exist to any appreciable extent; and there, as we have seen, they are abundant. When H. W. Fowler[7] expressed doubt as to whether there are any exact synonyms in English, he of course was referring only to ordinary language.

### *3. Closely related concepts should have similar names*

An obvious way to obtain similar names for related concepts is to incorporate the name of the basic concept in the names of those subsidiary to it. Thus, in referring to various forces, one would use, not "friction," "tension," "resistance," "reaction," . . . , but *frictional force, tensile force, resisting force, reacting force,* . . . . Other examples of a similar character will occur to the reader. The fact that the basic word incorporated in several terms may not have precisely the same operational definition in all of them is not objectionable, since the modifying word can be used to indicate the difference.

Another method of denoting concepts forming a closely related group is to use the same root words with different affixes and combining forms.[8]

[5] See *American standard definitions of electrical terms* (American Institute of Electrical Engineers, 1941). For each concept defined in this comprehensive glossary, the term considered the most acceptable is given in bold-face type and the synonyms, if any, in light-face type. Deprecated synonyms are so indicated by footnotes.

[6] See reference 2, p. 101; also, D. Roller, "The periodical literature of physics: . . . ," *Am. J. Physics* **14**, 300 (1946), esp. p. 306.

[7] H. W. Fowler, *A dictionary of modern English usage* (Oxford, 1930), p. 591.

[8] For a useful and enlightening discussion of this method, see Moon and Spencer, *Am. J. Physics* **14**, 285 (1946); also, reference 4.

The following affixes and combining forms are commonly used in physics with fairly definite meanings:

| | | |
|---|---|---|
| -ance | -lysis | -phorus |
| -chronous | -ment | -scope |
| -er | -meter | -scopy |
| -gram | -metry | -stat |
| -graph | -oid | -statics |
| -graphy | -on | -therm |
| -ity | -or | -tion |
| -kinetic | -osis | -tropic |
| -logy | -phone | |

| | | |
|---|---|---|
| a-, an- | helio-, heli- | mono-, mon- |
| allo- | hepta-, hept- | neo-, ne- |
| ana-, an- | hexa-, hex- | ondo- |
| anemo-, anem- | hydro-, hyd- | opto- |
| astro-, astr- | hygro-, hygr- | ortho-, orth- |
| atmo-, atm- | hyper- | pan-, pant-, panto- |
| baro-, bar- | hypo- | para-, par- |
| cata-, cat-, cath- | hypsi-, hyps-, hypso- | per- |
| centi- | | peri- |
| chrom-, chroma- | infra- | phono-, phon- |
| chromato-, chromat- | iso-, is- | phos- |
| chrono-, chron- | kilo- | photo- |
| cryo-, cry- | kineto- | piezo- |
| deci- | levo- | proto-, prot- |
| deka-, dek- | litho-, lith- | pyro-, pyr- |
| dextro- | macro-, macr- | recti-, rect- |
| dia-, di- | mano- | seismo- |
| endo-, end- | mega-, meg- | stereo-, ster- |
| exo-, ex- | meso-, mes- | syn-, sym- |
| geo- | meta-, met- | tachy-, tacho- |
| gonio- | metro- | tele-, tel- |
| gyro- | micro- | ter- |
| hecto-, hect- | milli- | tetra-, tetr- |
| helico-, helic- | myria-, myri- | thermo-, therm- |

A number of the foregoing affixes have, in truth, more than one meaning. But, as Moon and Spencer point out, complete consistency is not to be expected in any natural language. This does not, however, affect the desirability of associating definite scientific meanings with definite affixes.

If this practice of giving similar names to closely related concepts is to be effective, authors should avoid altering even slightly the name of a concept. Thus, if *index of refraction* is the accepted term, then "refractive index" should be avoided; and similarly for *friction coefficient* and "coefficient of friction," *thermal capacity* and "heat capacity," *potential difference* and "difference of potential," *radioactive element* and "radioelement," *viscosity coefficient* and "coefficient of viscosity" or "viscosity."

*4. Concepts not closely related should have names that differ markedly in appearance and sound*

Some pairs of terms that are unsatisfactory in this respect are: *absorption* and *adsorption; intermolecular* and *intramolecular; microscopic* and *macroscopic*. As C. G. Darwin[9] has said, any such pair of terms can confuse reader, printer, listener, and sometimes even the speaker, especially if they represent diametrically opposite ideas. *Microscopic* and *macroscopic* each consists of 11 letters "with only one of them different, and they are such that the pronunciation of either in some English dialects would give the impression that the other was meant." Darwin suggests *atomic* and *molar* as being better.

Confusion is likely to result if different, although perhaps cognate, concepts are assigned very similar names in an unsystematic fashion. For instance, all of the terms in any one of the following sets represent different ideas:

damping coefficient, d. constant, d. factor.
Debye temperature factor, D. characteristic temperature.
dichroism, dichromatism.
electric osmosis, electro-osmosis.
electrochemical equivalent, e. constant; faraday, farad.
Hamilton equations, H. principle, Hamiltonian, Hamiltonian function.
intensity of magnetism, magnetic intensity.

Some 15 terms representing different concepts contain the name "Maxwell;" and at least 12, the name "Lorentz."

*5. A term should be more or less self-explanatory*

Examples of terms that afford some suggestion of their technical meanings are *stress* and *strain; conductance, impedance, . . . ; retentivity, susceptibility, . . .* ; and so on. Many technical terms that have been coined from classical languages are somewhat self-explanatory because their component parts are words in the original language that have been selected with regard for meaningfulness. Many of the affixes listed under rule *3* are of material help in rendering more meaningful the terms in which they are incorporated.

On the other hand, some physical terms not only fail to explain; they actually foster misconceptions. Fortunately, some of these terms are now obsolete, or nearly so; examples are "black

[9] C. G. Darwin, "Terminology in physics," *Nature* **138**, 908 (1936).

radiation," "chemical rays," "dark heat," "gram atom," "gram molecule," "latent heat," "perfect gas," "permanent gas," "radiant heat," "specific heat of electricity." Other terms that may soon be obsolete are "atomic heat," "electric condenser," and "electric capacity." All of these terms have more satisfactory synonyms.

Agreement is not general as to what should be used to replace *electromotive force*. The general adoption of *emf* would be an improvement. Another possibility is "electromotance:"[10] but, as Moon and Spencer[11] point out, the suffix *-ance* should be reserved to denote a passive property of an entity—as in *capacitance, reluctance*, and so forth—whereas *emf* is an active concept—like *energy, power* and *current*. "Voltage" has been widely used, of course, and has advantages[12] as well as disadvantages[13]; and it does not fit in with Moon's suggestion[4] that the suffix *-age* be used to signify "per unit area."

There is another sort of term that is self-explanatory only if one is familiar with the history of the concept. This is the name adapted from that of the discoverer of the concept. Of all the terms in pure physics—including the special names given to generalizations, effects and processes, but not the names of units of measurement—at least 15 percent incorporate proper names.[14]

Among such terms, all of the following could be immediately discarded because each has a more meaningful or less ambiguous synonym:

Ångström coefficient, Aston dark space, Compton electron, Cotton-Mounton effect, Crookes dark space, Curie point, D'Arsonval instrument, Debye factor, Dorn effect, Edison effect, Faraday effect, Foucault currents, Fresnel zone, Haüy law, Hittorf numbers, Huygens zone, Lambert law, Lenard tube, Marx effect, Moseley number, Poinsot ellipsoid, Richardson effect, roentgenoscope, Ruhmkorff coil, Seebeck effect, Steinmetz coefficient, Verdet constant, Volta effect, Volta potential, Wimshurst machine.

The indication is not that all proper names for concepts necessarily should be discarded. Some concepts may be of such character that the proper name may be as good as any other, and may even be more important, for historical reasons. Moreover, terms incorporating proper names are likely to be international in character

[10] Proposed by A. W. Duff, *Am. J. Physics* **6**, 219 (1938).
[11] Reference 8, p. 289.
[12] A. Hazeltine, *Am. J. Physics* **15**, 191 (1947).
[13] D. Roller, *Am. J. Physics* **14**, 340 (1946).
[14] In Weld's *Glossary of physics* (McGraw-Hill, 1937), some 585 terms, or about 18 percent of the total, fall in this category. In reference 5, the policy apparently was to discourage the use of terms involving proper names.

(rule 7). As to objections to this method of naming concepts, an additional one is the wrong impression it fosters of how physical science has developed; in placing emphasis on the work of a very few individuals, important factors are left out of account, such as the socio-economic and, today especially, the fact that a great many individuals may have contributed to a particular discovery.

These objections do not appear to apply to the practice of honoring a great physicist by giving his name to a unit of measurement in his chief field of productivity. Such names—henry, joule, maxwell, newton, and so on—seem to be quite satisfactory for technical purposes, and they carry little implication that the physicist in question was solely responsible for some particular discovery.

Where proper names are retained, one certainly should make sure that the recognition is given to the real discoverer, as established by priority of publication. For instance, books on heat refer variously to Charles's law, Gay-Lussac's law, and the law of Charles and Gay-Lussac. But the truth seems to be that the volume-temperature law should be associated with the name of Gay-Lussac, and the pressure-temperature law with that of Amontons, who often is left unmentioned; and if the first of the two laws is to be given a double name, it should be, not Charles and Gay-Lussac, but Dalton and Gay-Lussac.[15] National pride seems frequently to have an effect on the name associated in different countries with the same discovery.

Although the trick of stringing together the names of several persons who have done the pioneer work on a concept yields a clumsy term, it is one way to evade the task of deciding strict priority. Even this device may fail: the term *Wentzel-Kramers-Brillouin method* sacrifices simplicity in order to give credit to several independent discoverers, yet omits the name of Jeffreys, another independent discoverer and apparently the one with the strictest priority.[9]

### *6. The name should be simple and euphonious*

Terms such as "quantity of charge" and "magnetic pole strength" not only lack simplicity but are redundant, the "quantity" and "strength" being superfluous. "Sinusoidal function" should give way to the simpler *sinoidal function*, a form that is sanctioned by dictionaries. "Coefficient of linear expansion" and "coefficient of volume expansion" are gradually being replaced by ***linear expansivity*** and ***volume expansivity***, which are not only simpler but utilize the convention that *-ity* signifies a passive property of a substance.

[15] W. J. Lyons, *Am. J. Physics* **6**, 256 (1938).

Mention should also be made of the undesirable tendency to form expressions in which a noun qualifies an adjective. Common examples are "gram-molecular weight," for which there already exists the preferable term *mole*, and "quantum theoretical methods." Darwin[9] has suggested that the adjective "quantal" be coined, and then *quantal methods* could be contrasted with *classical methods*.

Another step toward simplification in terminology would be the adoption of the rule that use generally be made of the simplest spellings of technical and semitechnical terms given in the text proper[16] of standard dictionaries. Among the spellings that might well be adopted immediately are:

| | | |
|---|---|---|
| algebraic | distil | percent |
| aline | draftsman | philosophic |
| analytic | esthetic | questionary |
| buret | fiber | sinoidal |
| caliper | focused | sulfur |
| calory | gage | technic |
| catalog | geometric | x-ray |
| disk | paraffin | |

Among the acceptable Anglicized plural forms are:

| | | |
|---|---|---|
| abscissas | criterions | minimums |
| antennas | formulas | momentums |
| apparatus | maximums | vacuums |
| appendixes | mediums | vertexes |

Certain of the older plural forms—*data*, *phenomena*, *quanta*, *spectra*—probably should be retained, at least for some time to come. Changes in language must be effected gradually, and hence usually without complete consistency, if any general acceptance of them is to be expected. Some authors of technical papers, although only a few, cling to older, more complicated spellings in a manner remindful of Lincoln's remark about his wife's relatives: "One *d* is enough for God, but the Todds had to have two."

"Technique," as commonly employed in physics, could easily be replaced without danger of ambiguity by the simpler and more easily pronounced "technic" (tĕk′nĭk). Some writers object to this change, for usage favors "technique" and a change in pronunciation also is involved. The resulting simplification, taken by itself, is small indeed; but it furnishes a particularly good example of how improvements in technical language can sometimes be legitimately effected by creating usage rather than merely by following it.[17] A few writers also prefer "alignment" to "alinement," although the former is less preferable, being a bad spelling of the French.

### 7. *The name preferably should have the same form in the chief languages of the world*

The arguments favoring internationality in the names of scientific concepts and practicable ways in which internationality can be accomplished have been so ably set forth by Moon and Spencer[18] that it is unnecessary to repeat them here. Of 210 scientific terms examined by these authors, 63 percent were found to have essentially the same form in English, German, French and Russian. Of the various suffixes in common use (see rule *3*), four were found to have a high degree of internationality. They are: *-ance* (=a passive property of an entity); *-or* (=a device); *-ity* (=a passive property of a substance); *-tion* (=a process). Moreover, many physical terms incorporating one or another of these four standardized suffixes were found to be either the same or easily recognizable in the chief Western languages. Thus internationality in terminology has already been realized to some extent, and the indications clearly are that extended efforts in this direction are by no means "visionary," "impracticable" or "unnecessary."

Often a concept originates in one country and is there given a name that cannot easily be incorporated in other languages without change. Instead of making literal translations into other languages, it is usually better to ignore the original term and to choose a new one—preferably one that will be acceptable internationally or, at least, that will best describe the idea in the language in question. An example of an attempt at literal translation is the term *zero-point energy*, which is ambiguous, clumsy and not even well translated.[9]

Darwin[9] gives another example of interest. Suppose that the alpha-particle had been discovered in Germany, so that the original name used to denote its range was "Reichsweite." If this had been translated literally in order to obtain the English term, we would today speak of the "reach-width," rather than of the *range*, of such particles.

[16] This excludes reformed spellings that have not come into general use.

[17] The comparable change from *unique* to "unic" is not contemplated; for one thing, "unic," unlike *technic*, is now obsolete

[18] P. Moon and D. E. Spencer, *Am. J. Physics* **14**, 285, 431 (1946); **15**, 84 (1947).

### *8. If an existing term is somewhat faulty, but firmly established, a brief historical or other explanatory comment should accompany its definition in glossaries*

Some very old terms naturally originated in misconceptions or in situations that, at the time, were anomalous or incompletely understood. No one probably would suggest discarding such well-established terms as *atom*, *atomic weight*, *hydrostatics* and *inertia*. But such terms warrant some special discussion in glossaries and textbooks. For instance, Maxwell,[19] in discussing *inertia*, advises the student to "read Faraday's essay on 'Mental Inertia' which will impress him with the proper metaphorical use of the phrase to express, not laziness, but habitude."

Other terms that may need special discussion are those that have popular as well as technical meanings. Examples are: absolute, critical, fundamental, normal, rate, scientific, specific, spontaneous, theoretical, uniform.

Troublesome pronunciations should be indicated in glossaries and even in textbooks, for one naturally tends to avoid using a word if he is uncertain of its pronunciation. Some words are mispronounced by omitting certain sounds—as is frequently done with *laboratory*, *temperature*, *vacuum;* others, by adding certain sounds—as with *circuit*, *conduit*, *tangential;* still others by syllabifying incorrectly—as with *buoyant*, *buret*, *chloride*. Misplaced accent is also common.

Foreign proper names are troublesome, and of course they are abundant in physics. The best rule apparently is to pronounce them, "as nearly as possible, as they are pronounced by the well-educated people of the different countries to which such names belong, with the exception of those very few celebrated names, such as . . . Galileo . . . , which may be said to have acquired an established English pronunciation."[20]

Words of these several types that are commonly mispronounced include:[21]

| | | |
|---|---|---|
| accurate | aurora | borealis |
| adiabatic | auxiliary | Tycho Brahe |
| adsorption | Bernoulli | buoyant |

[19] J. C. Maxwell, *Theory of heat* (Longmans, Green, ed. 9, 1888), p. 86.

[20] J. Thomas, *Universal pronouncing dictionary of biography* . . . (Lippincott, 1911).

[21] See also "The pronunciation of chemical words," a report of the Nomenclature, Spelling and Pronunciation Committee of the American Chemical Society, *Ind. Eng. Chem., news ed.* **12**, 202 (1934).

| | | |
|---|---|---|
| buret | goniometer | quantity |
| centrifugal | hundred | research |
| chloride | interstice | retort |
| circuit | isotropic | selenium |
| compass | Joule | sidereal |
| component | kilometer | tangential |
| conduit | kinetic | temperature |
| Dewar | laboratory | theory |
| equilibrate | lenses | titanium |
| experiment | mono- | tourmaline |
| exponent | penumbra | undulatory |
| fluorescence | piezo- | vacuum |
| Greenwich | Polaris | zero |

Many technical and semi-technical words are differently pronounced by different orthoëpists. Examples are: allotropy, apparatus, area, asymmetric, commutative, calorimetric, chloride, cyclic, data, distillate, lamellar, lever, molecule, nomenclature, quasi-, univalent. Both lē'vẽr and lĕv'ẽr are acceptable; also, mŏl'ē̇-kūl and mō'lē̇-kūl; and so on. Yet students and laymen must wonder why different physicists pronounce such terms differently and are unable to agree on the pronunciation of words in their technical vocabulary. Absolute standards of pronunciation may not be possible or even desirable, but there could be more uniformity than now exists.[22]

### Sources of New Terms

The question arises as to where to obtain a new term when all existing names for an old concept are faulty or when an entirely new concept is to be named.

(*1*) *Borrow the word from ordinary language.*—Many terms borrowed from ordinary English have proved to be quite satisfactory (see rule 5), especially when they are so chosen that their ordinary and technical uses will seldom be needed in the same context. On the other hand, such terms are likely to be ambiguous and, since they seldom can be taken over directly into other languages, will not acquire international status. Apparently this method of obtaining a new technical term should be used with considerable restraint.

(*2*) *Borrow a word from a living, foreign language.*—Gibbs did this when he took the French word *ensemble* and gave it a special meaning. Similarly, when the terms "characteristic function" and "characteristic energy" proved to be

[22] Reference 20 describes the method used in determining preferred pronunciations of chemical terms. Some 200 chemists and several orthoëpists participated in the study.

clumsy, we made a partial reversion to *eigenfunction* and *eigenenergy*. An objection to this general method is that the resulting term, while good for the borrower, may not be satisfactory for the language from which it was obtained. For instance, *ensemble* in French does not convey Gibbs's idea very well.[9]

Words taken from one language may be difficult to pronounce when incorporated in another. Thus, a foreign word introduced into English probably should be Anglicized in pronunciation and even spelling; this has already been done with such terms as *calory*, *fiber*, *gram*, *liter* and *meter*.

(*3*) *Adopt the name of the discoverer or inventor.*—This method, which was discussed in connection with rule *5*, is the easiest of all to apply. As a way to name units of measurement it appears to be excellent; but for most other concepts it should be a last resort, when other methods have failed to yield a satisfactory term.

(*4*) *Coin a term more or less arbitrarily.*—Coining terms by making an arbitrary although euphonious combination of letters of the alphabet might serve on occasions; but if there were many such terms, they would be hard to learn and continual resort to a technical dictionary would be necessary. Another possibility is to join the first letters of a phrase describing the concept, as in *radar* and *loran*. Other methods of coining are illustrated by *mho* and *cunico*, terms that are both meaningful and international.

(*5*) *Construct the term from parts taken from a classical language.*—This method of coining terms has important advantages. The term, being new, is immediately recognizable as technical. If its component parts are selected from fairly well-known words in the original language and with regard for meaningfulness, the term will also be to some extent self-explanatory.

Finally, the term can be so coined that it will be capable of adoption, without change, by all European languages. This is particularly true, according to Moon and Spencer,[23] when the source language is classical Greek. Of the 210 terms mentioned in connection with rule *7*, 82 percent of those of Greek origin have essentially the same form in English, French, German and Russian, whereas only 38 percent of those derived from other languages, including Latin, have international status.

[23] Reference 18, which should be consulted for details.

As was previously mentioned (rule *7*), a number of suffixes already have international standing, notably *-ance*, *-or*, *-ity* and *-tion*. The proposal[19] is that these be augmented by four new suffixes: *-os* (=basic concept); *-ent* (=per unit length); *-age* (=per unit area); *-um* (=per unit volume). Thus a set of basic Greek words together with these suffixes would yield a technical vocabulary that is both unique and international.

### AAPT Committee on Terminology

The Association's Committee on Terminology was set up in December 1936, with the writer as chairman. One member of the Committee, Harold K. Hughes, had previously done work on letter symbols, and he was made chairman of a subcommittee on symbols. Mainly because of Hughes's enthusiastic and painstaking efforts, the work of this group soon became so extensive that the Association in 1937 made it a separate committee. Its personnel at the same time accepted appointment from the American Standards Association as Subcommittee 10, on physics, of ASA Committee Z10 on Letter Symbols and Abbreviations for Use in Science and Engineering.

The work of this joint ASA-AAPT committee has been described in several reports.[24] The final report is now in type and will be issued as *Proposed American Standard Letter Symbols for Physics*. The members of the joint Committee are: H. K. Hughes, *chairman*, E. L. Chaffee, A. W. Foster, A. T. Jones, V. F. Lenzen, D. Roller, M. W. Zemansky.

Meanwhile the Association's Committee on Terminology had begun to function. The study initially was confined to terms considered to be highly faulty, and progress consequently was slow. With the approach of the war years, committee members gradually became absorbed in other work. During the war the committee ceased to function, except for some work done by the chairman on symbols (abbreviations) for the names of physical units.

Because of many changes affecting personnel during the war, it was decided in January 1946 to discharge the original Committee on Terminology and to form a new one, which now consists of H. K. Hughes, P. Moon, L. D. Weld, M. W. Zemansky and D. Roller, *chairman*. Additional members and consultants will be added as the work progresses.

Experience with the original committee indicated the need for a comprehensive program of study, and for a set of principles for selecting terms upon which committee members could agree. The writer therefore collected previously published information on terminology and, in the light of it, attempted to formulate suitable selection rules. The present paper is the result. Portions of it have appeared in digest form in the *Bulletin* (May 1946) of the Society for Applied Spectroscopy, in connection with the initiation by that society of a study of nomenclature in spectroscopy.

[24] Notably in *Am. J. Physics* **8**, 300 (1940). The list of symbols recommended has also appeared in recent editions of *Handbook of chemistry and physics* (Chemical Rubber Publishing Co.).

# Reproductions of Prints, Drawings and Paintings of Interest in the History of Physics

## 30. Egyptian and Assyrian Pictorial Representations of the "Mechanical Powers"

E. C. WATSON
*California Institute of Technology, Pasadena 4, California*

IN discussing the so-called "mechanical powers"—the lever, the wheel and axle, the pulley, the inclined plane, the wedge and the screw—many physics textbooks make the statement that rude implements of this kind have been found in ancient graves and that the Egyptian and Assyrian records contain pictorial representations of such appliances; but no textbook with which I am familiar gives references to authorities that might be consulted for more specific information regarding these devices, or attempts to support the bare statement of fact by reproducing photographs of such implements or their pictorial representations. Feeling the need for a modicum of this sort of material in my own lectures in order to arouse the students' interest and stimulate their historical sense, I addressed a request for information as to where it could be found to the Oriental Institute of the University of Chicago. The staff of the Oriental Institute very generously went to a great deal of trouble to collect material from various sources and sent me the photographs that are here reproduced. They also suggested the references that appear in the footnotes and at the end of this article and furnished most of the information for the text that follows. Since other teachers of physics may desire to make use of illustrative

PLATE 1. Relief in the tomb of Mereruka (*ca.* 2300 B.C.) at Sakkarah, Egypt, showing artisans using stone-drills. [Courtesy of the Oriental Institute of the University of Chicago.]

PLATE 2. Small copper model of a Sumerian war chariot of about 2800 B.C. [Courtesy of the Oriental Institute of the University of Chicago.]

PLATE 3. Reconstruction of the small copper model of a Sumerian war chariot shown in Plate 2. [Courtesy of the Oriental Institute of the University of Chicago.]

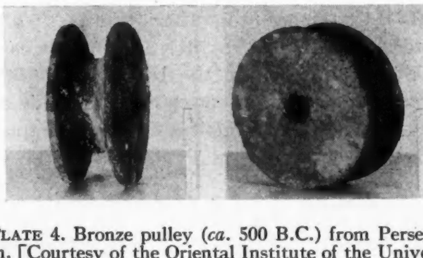


PLATE 4. Bronze pulley (*ca.* 500 B.C.) from Persepolis, Iran. [Courtesy of the Oriental Institute of the University of Chicago.]

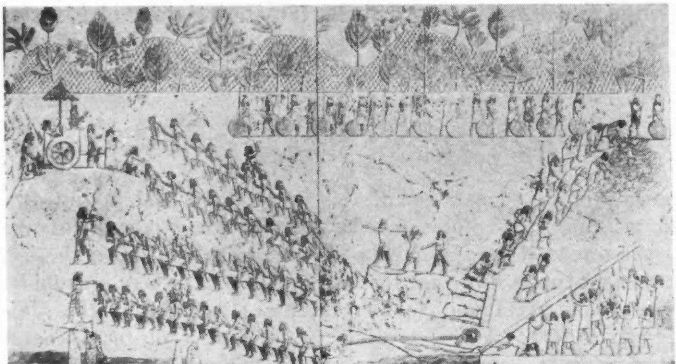


PLATE 5. Relief from the Palace of Sennacherib (705–681 B.C.) at Nineveh, Iraq. [Courtesy of the Oriental Institute of the University of Chicago.]

material of this kind, the Oriental Institute has most graciously granted permission for its publication as one of the numbers in this series of historical reproductions. My own best thanks, and I am sure those of many other teachers of physics, are due the staff of the Oriental Institute for their generosity both in assembling this information and in giving permission for its publication.

One of the first mechanical appliances of the Egyptians was the stone-drill (Plate 1) which was in use before 3000 B.C.[1] This consisted of a shaft with a handle, to the bottom of which was fastened a piece of flint. By rotating the shaft artisans could hollow out stone vessels.

Representations of the wheel with a horizontal axis, as a burden-bearing device for reducing friction, go back to about 2800 B.C. in Sumeria (modern Iraq) or about 1750 B.C. in Egypt. BREASTED states that "this type of wheel originated in Asia, presumably from round tree trunks used as rollers," and that "Sumerian chariots were known in Babylonia well toward 3000 B.C. Hence the chariot passed across western Asia to the Mediterranean and reached Egypt with Hyksos in the 18th century before Christ; but it was not common on the Nile until 1600 B.C. at the rise of the Egyptian empire."[2]

Plate 2 shows a small copper model of a Sumerian war chariot of about 2800 B.C. which was discovered by the Iraq expedition of the Oriental Institute at Tell Agrab, northwest of Baghdad, Iraq. The charioteer stands upright in the two-wheeled cart, which is drawn by four asses. A reconstruction, made by Messrs. Aumonier of London, is shown in Plate 3. The original model was only 3 in. high, but was made with great care and a high regard for detail. The rims of the chariot's wheels were "milled" with copper studs, probably to provide better traction, just as modern automobile and tractor tires have a tread. The model gives details of harnessing and explains how the charioteer kept his position on a springless vehicle. His feet rested on a small ledge above the axle, and his legs were astride a centerboard which he gripped with his knees. The yoke was fastened to the collars of the inner pair of animals, and the other pair of asses pulled on the collars of the inside pair. The reins were fastened to rings piercing the upper lips of the animals, and the driver's end was wrapped around the front part of the chariot. The wheels of the chariot are solid and made of three sections of wood clamped together. The reconstruction is made showing the charioteer with a whip, but he may have carried a spear instead; the corrosion of the original copper model makes this point uncertain.

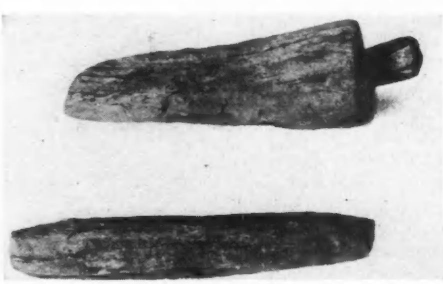


PLATE 6. Wedge and roller (*ca.* 2500 B.C.) from Sakkarah, Egypt. [Courtesy of the Oriental Institute of the University of Chicago.]

[1] J. H. Breasted, "Feats of old Egyptians rival modern works of engineers," *Popular Mechanics* 42, 403 (1924).

[2] See also Somers Clarke and R. Engelbach, *Ancient Egyptian masonry—the building craft* (Oxford Univ. Press, 1930), pp. 87–88, Figs. 82 and 83.

According to BREASTED[1] the pulley, another example of a wheel with a horizontal axis, may have been known to the Egyptians at an early date. However, the Oriental Institute informs me that alleged pulleys dating from before about 500 B.C. usually turn out to be metal or wooden guides, and SOMERS CLARKE and R. ENGELBACH[3] state that "the earliest known pulleys in Egypt are of Coptic or Roman date."

Plate 4 shows a bronze pulley from the terrace at Persepolis, Iran, which is now in the Oriental Institute. It dates from approximately 500 B.C. The Oriental Institute also has other pulleys from Egypt dating from 500 B.C. to the beginning of the Christian era. But, "while the pulleys used may have been a great convenience in changing the direction of the applied force without too great loss by friction, there is no evidence that they at the same time multiplied that force. With the exception of the lever, no device for multiplying force was known to the ancient oriental world."[1]

The use of the lever is excellently shown in a relief (Plate 5) from the Palace of Sennacherib (705–681 B.C.) at Nineveh, Iraq, which is now in the British Museum. This relief depicts the king in his chariot on a hill watching the transportation of a stone colossus on a sled on rollers (shown lengthwise in the picture and not crosswise as they were actually used) over a road prepared by pouring water on it. Slaves are using a lever to move the colossus onto the rollers, while other slaves pull ropes attached to the sled and three more slaves dip water out of the canal by means of the shaduf, an apparatus resembling a well-sweep (with bucket and counter weight). Further evidence for the early use of levers is given by CLARKE and ENGELBACH.[4]

[3] Reference 2, p. 44.
[4] Reference 2, pp. 88–89.

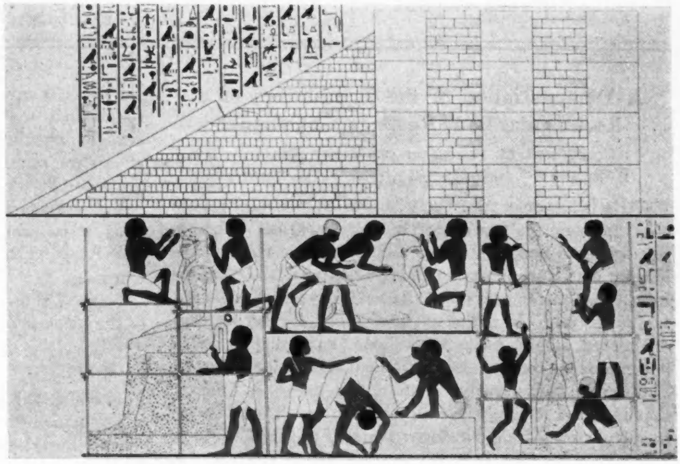


PLATE 7. Section of a mural painting in the tomb of Rekhmire (*ca.* 1501–1450 B.C.). [Courtesy of the Oriental Institute of the University of Chicago.]

Plate 6 shows a wedge, as used by stone masons for splitting stone, and a roller which were found near the pyramid of Pepi II, at Sakkarah, Egypt. They probably date from the 5th Dynasty (*ca.* 2500–2350 B.C.). The use of rollers in conjunction with sleds to reduce the friction is discussed in some detail by CLARKE and ENGELBACH.[5]

The inclined plane was commonly used by the Egyptians to raise blocks of stone to a desired height. Plate 7 reproduces a section of a mural painting in the tomb of Rekhmire (*ca.* 1501–1450 B.C.) at Thebes. In the upper half of the mural a ramp used in the construction of a temple is shown. In the lower half sculptors are depicted at their work. Other examples of constructional embankments are given by CLARKE and ENGELBACH.[6]

Additional information regarding early Egyptian and Assyrian mechanical appliances will be found in the references already cited and in *Die Technik des Altertums*, by ALBERT NEUBURGER (Leipzig, 1921).[7] The reader may also find it profitable to consult *Tools and Weapons*, by W. M. FLINDERS PETRIE (London, 1917).

[5] Reference 2, pp. 89–90.
[6] Reference 2, pp. 92–94.
[7] Translated by H. L. Brose under the title, *The technical arts and sciences of the ancients* (Macmillan, 1930); see especially pp. 202–218.

# NOTES AND DISCUSSION

## A Demonstration of the Equilibrium of a Rectangular Body Resting on a Cylinder

VIRGIL M. HUTCHISON AND HARRY HILL
*Washington and Jefferson College, Washington, Pennsylvania*

THIS demonstration was suggested by the classic problem in mechanics of showing that a heavy homogeneous body resting on a cylinder is in stable equilibrium if it is so placed on the cylinder that the distance from the center of gravity of the body to the point of contact is less than the radius of the cylinder. If both surfaces are curved, stability will exist so long as this distance does not exceed the product of the radii divided by their sum. If the radius of curvature of the upper surface is infinite, that is, if the surface is plane, this limiting distance is the radius of curvature of the lower surface.[1]

In this demonstration emphasis is placed on the curves representing the potential energies of rectangular blocks of various sizes when rocked on top of a cylinder of hardwood 6 in. in diameter and 3 in. long. These curves are traced on a sheet of paper fastened to a rigid upright board attached to the base on which the cylinder rests. The tracing pencil is held in a hole bored through the center of gravity of the block. When the distance from the center of gravity of the block to its point of contact on the cylinder exceeds the radius of the cylinder, the curve slopes downward from the original position when the block is rocked to one side of the balance point, indicating unstable equilibrium. When this distance is less than the radius, the slope of the curve is upward and there is stable equilibrium. It is found that the potential-energy curve is very nearly a straight line between certain limits when this distance equals the radius of the cylinder. The limits for this case are those at which the angle between the vertical radius and that through the point of contact cannot be considered equal to its own tangent as shown by the following equations. Beyond these points the downward slope of the curve increases.

Figure 1 shows that the height $y$ of the center of gravity of the block above the center of the cylinder is

$$y = r(2\cos\theta + \theta\sin\theta),$$

from which $dy/d\theta = r(\theta\cos\theta - \sin\theta) = 0$ when $\tan\theta = \theta$.

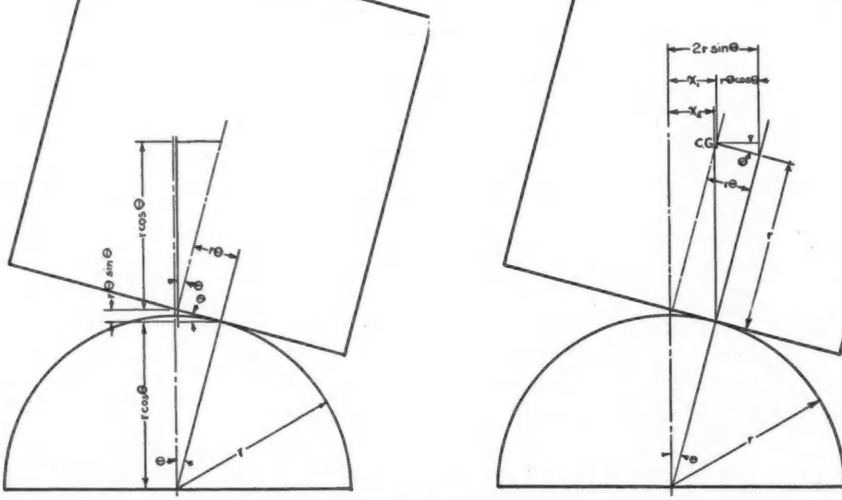
The same result (see Fig. 2) may be obtained by setting the horizontal distance of the center of gravity from the point of contact equal to zero. The distance of the center of gravity from the center of the cylinder is

$$x_1 = r(2\sin\theta - \theta\cos\theta).$$

The distance of the point of contact from the center of the cylinder is $x_2 = r\sin\theta$. Then

$$\begin{aligned} x_1 - x_2 &= r(2\sin\theta - \theta\cos\theta - \sin\theta) \\ &= r(\sin\theta - \theta\cos\theta) = 0 \end{aligned}$$

if $\sin\theta = \theta\cos\theta$, or $\tan\theta = \theta$. Hence, the equilibrium is approximately neutral for very small angular displacements.[2]




FIGS. 1 AND 2. Equilibrium of a rectangular body resting on a cylinder.

A very light plumb bob hung from the pencil shows the position of the vertical line through the center of gravity of the block relative to that through the point of contact and also to the one passing through the center of the cylinder. When the equilibrium is stable, the plumb line falls between these lines. In Fig. 2, the cylinder has already passed the point of stable equilibrium and the equilibrium is unstable.

[1] See Synge and Griffith, *Principles of mechanics* (McGraw-Hill, 1942), p. 218.

[2] See R. M. Sutton, ed., *Demonstration experiments in physics* (McGraw-Hill, 1938), Exp. M-40, for a similar demonstration in which the equilibrium is strictly neutral.

## Young Men in Physics

K. L. Yudowitch
*University of Missouri, Columbia, Missouri*

"WHEN were you at your mental prime?" It is the rare man indeed who will not answer: "I'm at my prime right now."

Hence, when recognition finally comes to the successful physicist and a biographer requests a picture, he invariably receives one of recent date. Of course, before the popularity of photography, it was only a recognized success who was in possession of a likeness of himself. But even today, the pictures that adorn our textbooks and propose to inspire our students are too often those of dignified elderly gentlemen.

The writer holds no prejudice against dignified elderly gentlemen; indeed, it is his ambition some day to become one. However, this custom of illustrating textbooks and biographies with likenesses taken in age has created a misconception in the minds of students The picture of a mature, bearded Galileo dropping objects from the Tower of Pisa belies the fact that his investigations on freely falling bodies were performed at the age of twenty-seven.

The effect of this false conception upon the attitude of the young student is very real. An age barrier is placed between him and the men whose works he studies. This barrier makes the material he reads less real to him. He finds little in common between his normal self and one whom he considers an aged man of abnormal mentality—a man who probably spent all his life in academic pursuits with no semblance of a normal existence.

The knowledge that so many important discoveries in physics have been made by young men comes as a surprise to most students—and a pleasant surprise. Students never fail to look with new interest upon work done by a man at very nearly their own age. Physics is revitalized in the minds of the students by the knowledge that it is a field for *young* men—men like themselves.

To counteract the students' false impressions, I have conveniently used marginal notes in the text, paying special heed to include men mentioned by name in the text for specific discoveries. The effort has been amply rewarded by increased class interest. Rearranged chronologically, a partial list follows:

| *Name* | *Contribution* | *Age* |
|---|---|---|
| Galileo | Simple pendulum | 19 |
| | Freely falling bodies | 27 |
| Snell | Snell's law | 30 |
| Pascal | Pascal's principle | 30 |
| Newton | Method of fluxions | 24 |
| | Law of gravitation | 24 |
| Young | Interference of light | 28 |
| Fresnel | Diffraction of light | 27 |
| Faraday | Electric motor | 30 |
| Kirchoff | Kirchoff's laws | 23 |
| Maxwell | Electromagnetic theory | 30 |
| Einstein | Restricted relativity | 26 |
| Bohr | Atomic theory | 28 |
| Compton | Compton effect | 30 |
| Heisenberg | Uncertainty principle | 24 |
| Van de Graaf | Van de Graaf generator | 30 |
| Anderson | Positron | 27 |
| Nier | Uranium isotope measurement | 28 |
| Kerst | Betatron | 30 |

## Is Not Voltage a Desirable Term?

Alan Hazeltine
*15 Tower Drive, Maplewood, New Jersey*

THE Note, "Misuse of the names of physical units," over the initials of the Editor,[1] is valuable in calling the attention of physics teachers to practices that tend to confuse the ideas of students. Exception may be taken, however, to one of the examples, the term *voltage*. Its status is perhaps best brought out in the following paragraph, written a half-century ago by an eminent physicist:[2]

> The line-integral of the electric force from one point to another along a stated path is the electromotive force along that path; this was abbreviated by Fleeming Jenkin to E.M.F. He was a practical man, as well as a practician. When expressed in terms of a certain unit called the volt, electromotive force may be, and often is, called the voltage. This is much better than "the volts." I think, however, that it may often be conveniently termed the voltage irrespective of any particular unit. We might put it this way. Volta was a distinguished man who made important researches connected with electromotive force, which is, therefore, called voltage, whilst a certain unit of voltage is called a volt. At any rate, we may try it and see how it works.

The practice here suggested by Heaviside, of using the term voltage irrespective of the unit in which it is ex-

pressed, has in fact become universal among workers with electrical apparatus. The term *electromotive force* has disappeared from their conversations and appears only occasionally in their publications, where it is felt to lend the dignity of a more ancient and legitimate ancestry. Such usage is particularly well illustrated by the host of compound phrases, such as voltage amplifier, voltage divider, voltage ratio, voltage regulator, voltage resonance, voltage transformer, and so on. No thought of the volt as a unit enters in these expressions; and never are the terms electromotive force or potential difference here substituted for *voltage*.

Is the association of the words voltage and volt really so damning? We have a similar association of the quantity *dioptry* and the unit *diopter;* and the term dioptry has been recommended (very properly) by the Editors in their "Information for contributors to the *American Journal of Physics*."

There need be no occasion for ambiguity in the term voltage. It is generic, signifying the quotient of any electric power by the accompanying current (existing or imagined), and is made definite by some qualifying adjective to indicate just which electric power is meant, where more than one power is present in a circuit element. Thus, for a voltaic battery, the *terminal* voltage corresponds to the output power and the *generated* voltage to the total power converted electrochemically. The use of such generic terms as voltage and power, with suitable qualifying adjectives, is far easier for the student than the use of unlike terms (such as electromotive force and potential difference) for different aspects of the same physical quanitity.

Rather, in respect to ambiguity, the shoe is on the other foot: the term electromotive force as found in the literature is ambiguous; for some writers use it generically, for any voltage, while others restrict it to generated voltages (that is, corresponding only to electric power which is not dissipated nor stored in a dielectric field). Sometimes these distinct usages are found in the same work. Such divergence is very confusing to students.

There are other objections to the term electromotive force: it is cumbersome, having six syllables and two words; and its inclusion of the word "force" misleads students and has to be explained away by teachers. The term voltage obviates all of these objections. Furthermore, there is always an advantage (though not necessarily a controlling one) in keeping the language of the classroom in accord with the language of the practitioner: it helps the student in his subsequent work and it increases his confidence in his teachers, to find them practical as well as theoretical.

The writer would advise habitual classroom use of the term voltage, dropping electromotive force entirely and employing potential difference only when the property of potential as a point function is a matter to be emphasized. This was his own practice over many years in teaching classes in electrical engineering[3] and later in physics.

[1] *Am. J. Physics* **14**, 340 (1946).
[2] O. Heaviside, *The Electrician* **26**, 507 (1891); reprinted in *Electromagnetic theory* **1**, 26.
[3] Hazeltine, *Electrical engineering* (Macmillan, 1924), esp. pp. 35, 36.

## Right Answer by Method Physically Wrong

Laurence Ellsworth Dodd
*University of California at Los Angeles, Los Angeles 24, California*

In a recent examination given to a large class of engineering students the following problem was included. Two cubical tanks, 3 ft and 18 in. on an edge, respectively, are connected by a pipe of length 2 ft and inside diameter 4 in., having a valve at its midpoint. The smaller side of the system contains nitrogen under a gage pressure of 20 lb/in.², and the larger side contains carbon dioxide under a gage pressure of 30 lb/in.² Find the resulting equilibrium pressure after the valve has been opened.

When the graded papers were returned, several members of the class asked why they were penalized on their solutions to this problem, since they had obtained the correct answer. It was pointed out that here they had encountered a problem in which a physically wrong method could be used to give a correct result.

The problem is of the well-known type involving both Boyle's law and the law of partial pressures, in which, of course, absolute pressures must be used. From Boyle's law,

$$p_2+B=(p_1+B)V_1/V_2,$$

where $p$ is gage pressure and $B$ is barometric pressure. Then

$$p_2=[p_1V_1+B(V_1-V_2)]/V_2.$$

Now from the law of partial pressures, $(P_2)_N+(P_2)_C=P_2$, where the subscripts $_N$ and $_C$ refer to nitrogen and carbon dioxide, respectively, and the $P_2$'s are absolute pressures. The final gage pressure in the present problem then becomes

$$\begin{aligned}p=P_2-B&=(P_2)_N+(P_2)_C-B\\&=\frac{(p_1)_N(V_1)_N+B[(V_1)_N-(V_2)_N]}{(V_2)_N}\\&\quad+\frac{(p_1)_C(V_1)_C+B[(V_1)_C-(V_2)_C]}{(V_2)_C}+B.\end{aligned}$$

But a condition of the problem is that

$$(V_1)_N+(V_1)_C=(V_2)_N=(V_2)_C=V_2.$$

Then

$$\begin{aligned}p&=\frac{(p_1)_N(V_1)_N+(p_1)_C(V_1)_C+BV_2-2BV_2}{V_2}+B\\&=\frac{(p_1)_N(V_1)_N+(p_1)_C(V_1)_C}{V_2}.\end{aligned}$$

The physically wrong method is as follows. From Boyle's law, by incorrect substitution,

$$p_2=p_1V_1/V_2,$$

where $p_2$ is a wrong value of the final pressure of a component gas; whence from the law of partial pressures, by using these wrong values,

$$\begin{aligned}p&=(p_2)_N+(p_2)_C\\&=(p_1)_N(V_1)_N/(V_2)_N+(p_1)_C(V_1)_C/(V_2)_C.\end{aligned}$$

But $(V_2)_N=(V_2)_C=V_2$. Hence

$$p=\frac{(p_1)_N(V_1)_N+(p_1)_C(V_1)_C}{V_2},$$

as before.

It may be argued that because the physically wrong method leads to the correct result it is only a mathematical short cut. The answer is that such a short cut is permissible provided the error in the physical method is clearly recognized. In the present case, the wrong method leads to the right result because of the terms in which the problem was stated, together with its particular physical conditions.

The situation may be summarized entirely algebraically, detached from physical meanings of the symbols. From simultaneous equations corresponding to (i) Boyle's law, (ii) the law of partial pressures, (iii) the relation between gage pressure and absolute pressure, and (iv) the particular condition of mixing that makes the final volumes of the two component gases equal to each other and also equal to the sum of their initial volumes, we arrive mathematically at the same solution by two different routes, one of them requiring less work than the other.

Several supplementary remarks of interest may be added.

(1) The same conclusion of right physical answer by wrong physical method applies also when three, or any number $n$, of component gases are mixed in the present manner by bringing into free communication the separate tanks in which the components are initially stored. With the right method, the final gage pressure is

$$p=[(P_2)_a+(P_2)_b+\cdots+(P_2)_n]-B,$$

which, after substitution from Boyle's law and with the condition that $(V_2)_a=(V_2)_b=\cdots=(V_2)_n=(V_2)$, may be put in the form

$$p=[\Sigma_a{}^n\,(p_1V_1)+B(\Sigma_a{}^n\,V_1-V_2)]/V_2.$$

This is subject to the condition $\Sigma\,V_1=V_2$, so that

$$p=\Sigma_a{}^n\,(p_1V_1)/V_2.$$

With the wrong method,

$$p=(p_2)_a+(p_2)_b+\cdots+(p_2)_n,$$

where the $p_2$'s are wrong partial pressures. Substituting, wrongly, from Boyle's law, this becomes the correct answer, as previously given.

(2) With $n$ components, the number of arithmetical steps saved in calculating by the method physically wrong is $(n+1)$, the total number of steps by the method physically right being $(3n+2)$ and that by the method physically wrong being $(2n+1)$.

(3) Regardless of the manner of mixing three or more component gases, whether simultaneously for all components or by first mixing two or more of them together, then adding this mixture to the remaining components—that is, regardless of the combinations used in the mixing process—a right final answer always results by use of the physically wrong method of calculation. For example, and symbolically, with three components, $a$, $b$, $c$, we may indicate simultaneous mixing by $a+b+c$, and other possible ways by $(a+b)+c$, $a+(b+c)$, and $(c+a)+b$. Correspondingly, indicating the right final answer by $R$, and a physically wrong value by $W$, we have for the first case, $R=W_a+W_b+W_c$, and for the others, $R=(W_a+W_b)+W_c=R_{ab}+W_c$, and so on. In brief, when any number $n$ of components, no matter how many values, whether all of them are wrong or some of them are right and the rest wrong, are added up, the sum is always correct.

## A Type of Equipment Useful in Teaching Electronics

Robert Stollberg
*Wabash College, Crawfordsville, Indiana*

THE equipment described in this note consists of a selection of electronics components so designed that they can be arranged into any common circuit, as in Fig. 1. Each unit is mounted on an individual plaque, as in Fig. 2, and equipped with snap fasteners to provide electrical interconnections among parts. The plaques are drilled with a square pattern of holes which can be slipped over a similarly spaced array of nails or screws in a vertical mounting panel. Thus the units can be spread out on a table top for study by individuals or small groups, or can be displayed on an upright panel for classroom demonstration.

The most common size of plaque, such as that bearing a tube socket, is 3½ in. square and ¾ in. thick. This size, along with the 2-in. spacing of nails, permits most small electronics components to be supported without difficulty. An assembly panel 4×8 ft is found to be none too large for classroom demonstration, although one considerably smaller is suitable for laboratory experiments. Plaques are made of straight grained hard wood, while ¾-in. plywood provides an excellent mounting panel.

Smaller plaques, shown in Fig. 3, are employed for resistors, capacitors and other parts whose number and variety prohibit an individual mount for each possible component; the plaques are equipped with snap fasteners

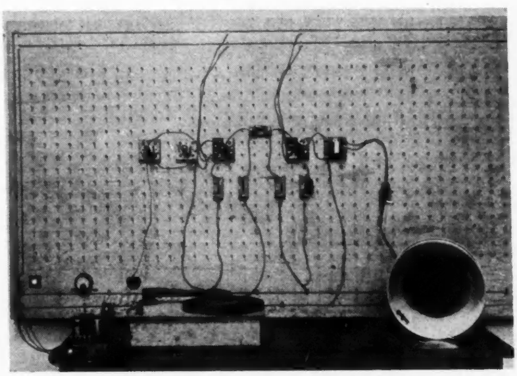


Fig. 1. Simple but complete audio amplifier mounted on assembly panel.

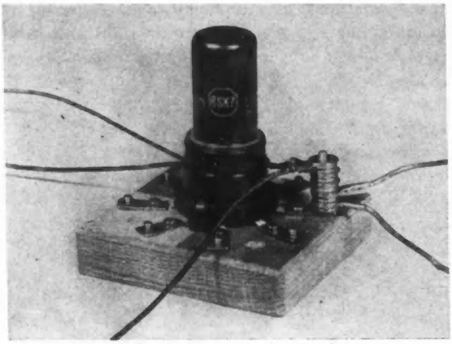


FIG. 2. Typical plaque showing mounting holes, color coded socket terminals, snap fasteners and connecting wires.

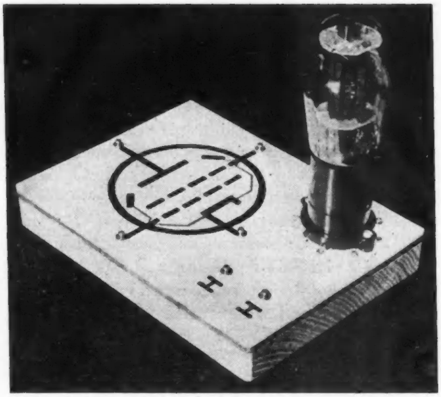


FIG. 4. Combination component-symbol plaque. Tube elements are invisibly wired to snaps on symbol. Tube heater terminals are available at snaps $H$, $H$.

for connecting wires and with Fahnstock clips for holding the desired parts.

Some of the distinctive and useful features of this equipment are:

(1) Each component is on an individual plaque, and so can be incorporated into any circuit; moreover, the parts can be oriented so as to resemble the corresponding circuit diagram.

(2) Tube socket terminals (Fig. 2) are identified with the standard RMA color code. The same is true of other terminals that are ordinarily numbered, such as the connections of a universal output transformer.

(3) For advanced work, components are mounted on plain plaques; but for elementary instruction, large facsimiles of standard symbols can be affixed to the plaques along with the electronic components themselves. Connections in this case can be made at the symbol rather than at the part, with concealed wiring connecting corresponding parts of symbol and component. Such a plaque is shown in Fig. 4. In the case of tubes, miniature battery-operated models can be employed, with cell, switch and tube integral with the plaque.

(4) To provide electrical interconnections, all plaques are equipped with male snap connectors, while assorted lengths of flexible insulated wire are equipped with combination snaps. These are female on the bottom, male on top, thus permitting indefinite piling up of connectors at a single junction (Fig. 2). Modifications of connecting wires permit connections to grid caps, shielding of wires, and so forth.

(5) The assembly panel is equipped with two filament-heater buses at the top and with a ground bus and a $B$-plus bus at the bottom. These are color coded and generously equipped with male snap fasteners.

(6) When a circuit is mounted on the panel, the network is fed by a separate power supply of conventional design, except that a bleeder resistor across the high-voltage terminals is equipped with a variable tap. When the negative terminal of the power supply is grounded, the tap can be used for low-voltage requirements, screen supplies, and so forth. When the tap is grounded, the negative terminal of the power supply becomes available for bias voltages.

The only real limit to the construction possibilities of such equipment is that imposed by high frequency. Actually, the broadcast band can be worked quite readily, with only moderate care in shielding. All kinds of components can be employed, including tubes, resistors, capacitors, inductors, transformers, speakers, potentiometers, switches, fuses and meter movements. Simple lights, motors, switches, fuses and so on would make this type of equipment equally useful in the study of elementary electricity.

Typical circuits which can be constructed from this equipment are: power supplies, audio amplifiers, tuned-radio-frequency receivers, superheterodyne receivers, low-power transmitters, test and indicating equipment, audio and radio frequency oscillators, cathode-ray oscilloscopes and complex networks.

In addition to the applications in classroom demonstrations and laboratory experiments, this equipment finds use as a core for group discussion and cooperative construction activities, a medium for trouble-shooting experiences, an aid to electronics circuit design, and a basis for a new technic of student evaluation in electronics.

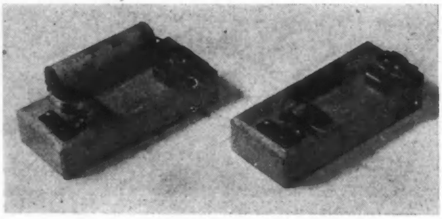


FIG. 3. General utility plaques for small components.

# Proceedings of the American Association of Physics Teachers

## The Sixteenth Annual Meeting, New York, January 30, 31 and February 1, 1947

THE sixteenth annual meeting of the American Association of Physics Teachers was held at Columbia University, New York City, on January 30, 31 and February 1, 1947. The presiding officers were R. C. Gibbs, President of the Association, and Paul Kirkpatrick, Vice President. Professor Kirkpatrick was chairman of the program committee for the meeting.

A joint dinner with the American Physical Society was held on Friday evening, January 30. Because of the large number of reservations, it was necessary to serve the dinner in both the Men's Faculty Club and John Jay Hall.

### Invited Papers

*Joint Symposium with the American Physical Society: Physics and National Affairs*

**Research and development in the War Department,** H. S. AURAND, *Director of Research and Development, War Department.*

**Research programs of the Office of Naval Research in the physical sciences.** P. F. LEE, *Chief of Naval Research, Navy Department.*

*Joint Sessions with the American Physical Society*

**Presentation of the Oersted Medal of the American Association of Physics Teachers to Duane Roller.** L. W. TAYLOR, *Chairman of the Committee on Awards,* and R. C. GIBBS, *President of the Association.*

**An approach to the study of physical terminology,** DUANE ROLLER, *Wabash College.*

**The Franck-Condon principle—address of the Retiring President of the American Physical Society.** E. U. CONDON, *The National Bureau of Standards.*

**The present status of atomic physics—fifth Richtmyer Memorial Lecture of the American Association of Physics Teachers.** J. R. OPPENHEIMER, *University of California.*

**The role of the National Research Council in American science and education,** D. W. BRONK, *Chairman of the National Research Council.*

*Symposium: Physics for Students of Biology and Medicine*

**Objectives of the Association's committee for students of biology and medicine.** LEROY L. BARNES, *Cornell University.*

**The role of biology in the physics curriculum,** D. W. BRONK, *University of Pennsylvania.*

**Sample illustrations of fundamental physical principles selected from biology and medicine.** LOUIS A. STRAIT, *University of California.*

**A course in physical instrumentation for medical and biological research.** K. S. LION and F. O. SCHMITT, *Massachusetts Institute of Technology.*

### Contributed Papers, With Abstracts

Two sessions were devoted to the following contributed papers:

**1. Standing-wave demonstrations.** M. J. PRYOR, *New York State College for Teachers.*—Longitudinal standing waves are set up in a helical spring in such a manner as to make them as vivid as the transverse standing waves produced in a string. Both are shown simultaneously, if desired, by attaching both spring and string to an eccentric of the type devised by Wold and Studer.[1] The similarity in method of demonstration leads directly to the common explanation. The explanation is made convincing by means of an apparatus[2] that consists of two rotating cylinders with sine curves cut in one edge of each. The apparent displacements in each of two oppositely directed continuous waves are graphically added in such a manner that an indicator disturbed by both waves is left at rest while another indicator subjected to the same displacements as the first is made to move with considerable amplitude. The separation of these indicators is $\frac{1}{4}\lambda$, as is the separation of a node and antinode in the spring and string that demonstrate standing waves.

[1] Wold and Studer, *Am. J. Physics* **8**, 165–167 (1940).
[2] Pryor, *Am. J. Physics* **13**, 110–111 (1945).

**2. Elementary derivation of mass-energy relation.** J. G. WINANS, *University of Wisconsin.*—Length $s$, time $t$ and force $F$ are taken as fundamental undefined quantities; velocity $v[=ds/dt]$, acceleration $a[=dv/dt]$, impulse $I[=\int Fdt]$ and work $w[=\int Fds]$ are defined in terms of these fundamental quantities; energy is defined as stored work. Experiments show that, for objects starting from rest, $(\int Fdt)/v=$const., and we may call this constant $m$, or mass. Then since $m=m_0(1-\beta^2)^{-\frac{1}{2}}$, where $\beta=v/c$, we have $\int Fdt=m_0\beta c(1-\beta^2)^{-\frac{1}{2}}$. Differentiation and then multiplication by $ds/dt[=\beta c]$ gives $Fds=m_0c^2\beta d\beta(1-\beta^2)^{-\frac{3}{2}}$. Integration then yields $\int Fds=m_0c^2(1-\beta^2)^{-\frac{1}{2}}-m_0c^2$, or $w=(m-m_0)c^2$.

**3. Dynamical problems in the evolution of the solar system,** HERBERT JEHLE, *The Franklin Institute.*—Theories of the origin of the solar system from a collision or close encounter between stars suggest that planets and their satellites were formed at the time of the encounter. They cannot explain the present outstanding regularities in the

solar system, such as small eccentricities and inclinations of orbits, absence of retrograde particles and conspicuous relationships between orbital elements, particularly distribution of mean motions. The formation of the solar system from a chaotic swarm of smaller particles, however, offers a possibility of explaining the aforementioned features. It is assumed that this swarm, because of its accidental resultant angular momentum, developed into a flattened disk-shaped system and that retrograde particles were eliminated because of their interaction with a majority of directly revolving particles. A chance accumulation of many particles toward orbital elements of perhaps Jupiter and Saturn caused other particles to tend toward preferential orbits and then agglomerate into larger bodies, thereby accounting for the location of planets and satellites. The dynamical problems involved in describing these processes lead to self-consistent transient solutions of the hydrodynamical equations and are typical of restricted problems of four bodies. Weizsaecker's dynamical approach based on the two-body problem is too primitive.

**4. On a generalization of Poiseuille's law.** ALLEN L. KING, *Dartmouth College.*—Poiseuille's law is generalized for the steady flow of fluids in cylindrical tubes with distensible walls. The volume flowing per unit time through a cross section of radius $r$ is given by

$$Q=\pi r^2\int_0^{v'}\eta v\,dv\Big/\int_0^{v'}\eta\,dv,$$

where $v'$ is the fluid velocity at the center of the tube. The coefficient of viscosity $\eta$ may be a function of velocity and velocity gradient. This result is applied to the flow of fluids of constant viscosity through tubes (i) with rigid walls, (ii) with thin walls of a Hookian material, and (iii) with thin elastomeric walls.

**5. Demonstration of Kirchhoff's law of radiation.** MARIO IONA, JR., *University of Denver.*—The experiment described shows that transparent objects radiate much less than other bodies at the same temperature. A platinum wire is heated inside of a quartz tube by means of a torch. The wire appears much brighter than the surrounding transparent quartz tube, which is certainly not at a lower temperature. If a glass tube is used, care must be taken that the glass does not flow too much during the experiment, and filters should be interposed to eliminate the bright light of the sodium flame. The wire should preferably have a characteristic shape—for example, a helix—so that it cannot be mistaken for the surrounding tube. An oxygen torch causes the wire to be bright enough for projection of its image on a screen.

**6. That trouble maker—the factor two.** W. H. MICHENER, *Carnegie Institute of Technology.*—Several problems, when analyzed by two different methods equally plausible to the student, yield results that differ by a factor of two. (i) Sand pours at a constant rate onto a conveyer belt moving horizontally; calculate the force required to keep the belt moving with a constant velocity. Calculation of the force from the time-rate of change of momentum yields a horizontal force twice as large as that calculated to give the sand its kinetic energy. (ii) Water escapes from a hole in the side of a tank. Energy relations give the familiar equation $v^2=2gh$; considerations of momentum, with plausible assumptions, yield $v^2=gh$. (iii) A battery, a switch and a capacitor are connected in series. When the switch is closed, the battery delivers the energy $qV$; the final energy acquired by the capacitor is $\frac{1}{2}qV$. Even though these problems may at first appear unrelated, their explanation involves the same principles. They make one more cautious in applying the principle of conservation of energy.

**7. Similarities of magnetic circuits and incandescent lamps.** RICHARD C. HITCHCOCK, *State Teachers College, Indiana, Pennsylvania.* Ferromagnetic permeability varies with flux. A similar electric situation is an incandescent filament, the resistivity of which varies with temperature. As an introduction to magnetic-circuit problems, it is suggested that a few problems in electricity be given, utilizing the more familiar electric quantities. The following data are measured for each of several standard 120-v tungsten lamps: $I$ (amp), $A$ (cm²); $I/A$ (amp/cm²), $E$ (v), $l$ (cm) and $E/l$ (v/cm). An experimental curve of $I/A$ as a function of $E/l$ is plotted for each lamp. The cross-sectional area $A$ and the stretched-out length $l$ of the filament were supplied by the manufacturer. The corresponding magnetic data are: $\Phi$ (maxwell), $A$ (cm²), $\Phi/A[=B]$ (maxwell/cm²=gauss), $F$ (gilbert), $l$ (cm) and $F/l[=H]$ (gilbert/cm=oersted).

The graph of $B$ as a function of $H$ is similar to those obtained in the electrical case. In the electric circuit, $E$ and $I$ are related by the equation $R$ (ohm) $=E$ (v)$/I$ (amp) $=l\rho/A$. Solution of this equation for the resistivity $\rho$ yields $\rho=(E/l)/(I/A)$[(v/cm)/(amp/cm²)=ohm/cm[illegible]]. For the magnetic circuit, the relation is $R=F\Phi$ (gilbert/maxwell) $=l/\mu A$. Solving for the permeability $\mu$, one gets $\mu=(\Phi/A)/(F/l)=B/H$ (gauss/oersted).

**8. Experiments to demonstrate the Mach law of inertia.** AUSTIN J. O'LEARY, *The City College of New York.*—Described in part in *Am. J. Physics* **15**, 146 (1947); also, *Am. J. Physics* **14**, 120–123 (1946).

**9. Statistical survey of numbers of physicists in training.** MARSH W. WHITE, *The Pennsylvania State College.*—The shortage of physics teachers and the large number of applicants for admission to the physics curriculum in recognized graduate schools made it desirable that a statistical survey be undertaken of the numbers of prospective physicists currently in training for degrees. Following conferences with some of those who need these data for establishing long-range policies on the training and utilization of physicists, it was agreed that Sigma Pi Sigma, physics honor society, should make a survey of the numbers of student physicists currently in educational institutions in this country. A postal-card questionary was sent to all departments that are recognized as offering a physics major. Replies received from the majority of these departments have been analyzed. The resulting data show: (i) the

numbers of B.S. candidates in each of the current four collegiate years, (ii) the numbers of M.S. candidates expected to be graduated in 1946–47 and other years, and (iii) the numbers of Ph.D. degrees likely to be granted in the same period. The data show a large increase over the enrolments of prewar years. Unless a better distribution of graduate students is made than is currently indicated, it is probable that the more familiar institutions will be swamped with candidates while many excellent institutions may have facilities to accommodate more students than will apply.

**10. Two new wave models.** HAROLD K. SCHILLING, *The Pennsylvania State College.*—Two new wave machines are described. The first, which is hand driven, demonstrates three types of stationary wave simultaneously: transverse, longitudinal, and "water." The "particles" of the different wave types are interconnected so that all of them are caused to move by turning one handle; this is analogous to the well-known apparatus illustrating progressing waves of the same three types. The second machine, which is motor driven, is a simple wave synthesizer that illustrates the superposition, or "addition," of two waves going either in the same or in opposite directions, and at the same or different velocities. Both component waves, as well as the resultant wave, are visible at all times, so that it becomes clear how the process of "addition" proceeds.

**11. An objective basis for reorganizing physics teaching.** ROGERS D. RUSK, *Mount Holyoke College.*—Of the many standards of evaluation of physics teaching that may be set up, some are subjective and some objective. Those that deal chiefly with the student response, such as awakening of student interest and intellectual satisfaction in solving a problem, are subjective. Those that deal more specifically with problems as they exist in the external world, the present state of science, the tools and instruments most useful today, and the situations most frequently encountered or expected to be encountered, are objective. Both are ultimately aimed at best adjusting the student to his environment and making him useful to society and valuable to himself. An objective method of evaluation is here suggested in which, to take the laboratory course in general physics as an example, each experiment is broken down into its component elements. The attempt is then made to judge these in terms of frequency of use, frequency of application in life itself and relatedness to other component elements. For example, in 40 experiments that were surveyed, the meter stick was by far the most frequently used tool. Other tools that seemed of equal value were met only a very few times and were poorly related to the body of the course. Mental operations may likewise be judged. While no such analysis can be final, it does seem to throw many things into perspective whereby the general physics course may be better evaluated.

**12. Condensers in a.c. and d.c. circuits.** W. B. PIETENPOL, *University of Colorado.*—By placing a condenser of relatively small capacitance in series with a small flashlight lamp, the displacement current when 110-v a.c. is applied can be shown effectively. This is the conventional method, but it is pointed out that inexpensive equipment may be used. With the same arrangement and same value of capacitance, but a more sensitive lamp, a brilliant flash is produced when the circuit is closed using 110-v d.c. Opening and again closing the switch produces no effect unless the condenser is first discharged. Furthermore, the effect of line resistance in determining the rate at which the condenser is charged is shown. With capacitance and lamp properly selected, no light is produced on closing the switch, but discharging the condenser through the lamp produces a flash. The equipment is very satisfactory for showing series and parallel arrangement of condensers and for pointing out that when alternating current is used the condenser may retain a charge, the magnitude of which depends upon the instant at which the circuit is opened.

**13. Treatment of energy interchanges in elementary teaching.** ERIC M. ROGERS, *Princeton University.*—How far does the beginner understand the foundations of our belief in conservation of energy? Does he realize that all our energy studies deal with *interchanges* of energy between one form and another? To help in studies of energy interchanges, some viewpoints and demonstration experiments are suggested for treatment of *work*, *electric potential difference* and *emf*.

**14. A basic-concepts course in physics.** A. G. WORTHING, *University of Pittsburgh.*—This course for senior and early graduate students emphasizes the bases of our physical concepts. Through discussions and problems, it seeks to establish certain principles more firmly in the student's mind and to produce a unified consistent view of physics as a whole. Precision of statement is insisted on in the knowledge that slight peculiarities of speech frequently result from erroneous ideas. Further, students are encouraged to question procedures that are not clear. Definitions found in physics textbooks or developed by the students themselves are scrutinized on the basis that a definition of a term is a statement in more basic or simple terms which may replace the term defined without any change of thought. The policy of interpreting the symbols of equations as physical quantities rather than numerics is followed. The simple procedure of applying conversion factors in problems to satisfy individual preferences for units is rather highly developed. The course has grown from a two-semester, one-credit per semester course covering the whole of physics to a two-semester, three-credits per semester course covering mechanics only. Originally the student personnel consisted of the department's teaching assistants. Now, the enrolment, several times as large, includes many nonphysics majors.

**15. Remarks on an operational analysis of the concepts of force and of mass.** GERALD J. HOLTON, *Harvard University.*—This is an attempt at an operational analysis in terms of generalized systems such as rigid constellations. The definition of the force concept is seen to be governed by the mathematical convenience of postulating the superposition theorem. Contrary to historical procedure, the

principle of sufficient reason is not invoked explicitly. The concept of mass is considered to have its primary qualitative significance in defining a universal attribute of all arbitrary systems, rather than in being the consequence of either the definition of force, or observations in any specific single system. Here the law of action and reaction need not be postulated prior to the definition of mass, but is regarded as its experimentally observable consequence.

**16. Two experiments for the sophomore physics laboratory.** G. P. BREWINGTON, *Lawrence Institute of Technology.*—A convenient apparatus, which has proved quite satisfactory for sophomore laboratory experiments in viscosity of liquids, is described. Means are provided for making all measurements required in the Hagen-Poiseuille equation, and design changes are suggested to make the apparatus inexpensive, as well as to minimize the quantity of water usually spilled by the average sophomore. A second experiment, well-suited for the substantial introductory laboratory course, is designed about model theory of electric circuits. A model in which all resistances are equal to those in an actual circuit is described. If a potential difference 0.01 that applied to the actual circuit is applied to the model, the *ir* drop in each resistance is 0.01 and the power expended is 0.0001 that in the actual circuit; the potentials are adequately measured by the ordinary 2-m form of potentiometer. This apparatus is sufficiently flexible to engage the attention of better students for a complete laboratory period. It is of interest to note that large power networks are now being studied by similar models.

**17. Experiments on thermionic emission.** GERALD J. HOLTON, *Harvard University.*—Relatively simple apparatus has been used to develop a set of experiments on several important aspects of thermionic emission for our laboratory course in Atomic Physics. The presentation of similar experiments in standard textbooks often suffers from ambiguous or even erroneous derivations of formulas involved. A brief discussion of such flaws is followed by an account of experimental and theoretical equipment needed for the following exercises: verification of the Maxwell velocity distribution of emitted electrons; determination of the contact potential and the temperature of a heated cathode; verification of Child's law and Richardson's equation; determination of the work function of a filament; illustration of the Schottky effect.

**18. A laboratory experiment involving uncertainties of measurement.** O. H. BLACKWOOD, *University of Pittsburgh.*—We teach the determination of uncertainties with the aid of an experiment in which the student measures the length, width and thickness of a metal strip ($18\times1.2\times0.2$ cm). He measures its length several times by means of a meter stick and computes the mean length; then he finds the mean deviation, or "uncertainty," and the percentage uncertainty of the determination. He performs the same operations for the width of the strip, using vernier calipers, and for the thickness, using micrometer calipers. Using the mean values of the three dimensions, he computes the volume of the strip, keeping the proper number of significant digits *at each step* of the computation. He finds the percentage uncertainty (not a standardized precision index) of the volume measurement by adding the three percentage uncertainties. Students find this experiment more interesting than the one formerly used in which the dimensions of a large block were all found by means of a meter stick. Instructors think highly of it.

**19. Origin of successful doctoral candidates.** M. H. TRYTTEN, *National Research Council.*—The Office of Scientific Personnel has been making a study under a grant from the Carnegie Corporation to determine the origin of successful doctoral candidates. Information is now nearly complete on persons who have been granted the doctorate in the years 1936–45, inclusive, giving the undergraduate institutions at which they took their first academic degrees. From this information some interesting indications with regard to productivity of different types of institution furnishing doctoral candidates is available. Some preliminary analysis is given.

**20. Design and use of an x-ray spectroscope.** PAUL KIRKPATRICK, *Stanford University.*—The x-ray spectroscope[1] has been improved in design, and its use in instruction has been extended. In this instrument Bragg reflections are observed in a telescope eyepiece with a disk of fluorescent screen at the focal plane. With a 1-kw Mo tube the sensitivity and precision are ample for the following: observation of $K$ lines and wavelength measurement in several orders; resolution of $K\alpha_1\alpha_2$ with calcite; observation of relative intensities of lines; demonstration of increase of resolution with order; observation of continuous spectrum and measurement of short-wave limit with determination of $h/e$; measurement of grating spaces of various crystals and of variation of grating space with plane indices; observation of variation of reflected intensities with order; demonstration of effect of grinding and polishing upon reflection; measurement of wavelengths of absorption discontinuities (Moseley graph data); qualitative observation of variation of absorption with wavelength; illustration of monochromatizing effect of zirconium filter; demonstration of nondependence of absorption limits on chemical combination; identification of unknowns by absorption-limit observations, observation of total (noncrystalline) reflection from glasses, crystals and metals; and determination of x-ray refractive indices and the value of $e/m$ therefrom.

[1] P. Kirkpatrick, *Am. J. Physics* **10**, 233 (1942).

**21. A model atomic bomb.** RICHARD M. SUTTON, *Haverford College.*—To appear in full in a future issue of this journal.

---

*(The annual reports of the Secretary and Treasurer will appear in the next issue.)*